

MARIA AUGUSTA SCHMIDT

**Obtenção e caracterização de uma estirpe mutante no gene *fixN*
de *Herbaspirillum seropedicae***

Monografia apresentada como requisito
parcial à obtenção do título de Bacharel em
Ciências Biológicas, Setor de Ciências
Biológicas, Universidade Federal do Paraná.

MONBIO
SCHMIDT, Maria Augusta

Curitiba

2006

Orientadores:

Prof^a Dra. Rose Adele Monteiro

Prof^o Dr. Emanuel Maltempi de Souza

À Deus e À minha família.

AGRADECIMENTOS

Agradeço primeiramente ao Professor Fábio de Oliveira Pedrosa pela oportunidade de trabalhar e desenvolver minha Monografia no Núcleo de Fixação Biológica de Nitrogênio, pelas idéias, sugestões e todo interesse demonstrado.

Agradeço imensamente ao Professor Emanuel Maltempi de Souza, por ter aparecido e acrescentado em minha vida. Por todas as demonstrações de carinho e preocupação. Pelas grandes idéias, pois é uma das pessoas mais geniais que conheço, por todo interesse, dedicação, tempo, paciência e alegria que me ofereceu. Muito obrigada.

A todos os professores do laboratório, Prof^a Maria Berenice, Prof^a Roseli, Prof^o Leonardo, Prof^a Liu, Prof^a Leda e Prof^o Geraldo, pela amizade, sugestões e conselhos.

À Rô da coordenação do Curso de Ciências Biológicas, pela assistência e pela paciência. Também ao Seu Sebastião, pela alegria e incentivo.

Aos funcionários do Departamento de Biquímica, Dona Julieta, Dona Roseli Prado, Dona Nanci, Dona Marilza por toda dedicação e disposição dispensados e especialmente ao meu amigo Valter de Baura, por todos os favores, conselhos, idéias, colaborações, músicas, piadas, risadas enfim, por todos os momentos.

Gostaria de agradecer aos meus amigos queridos que serão para sempre especiais para mim Hudson, Eduardo, Michelle, Marco Aurélio, Marco Antônio, Dani e Ane. Obrigada pela confiança, convivência e carinho oferecidos.

Aos amigos do laboratório de Fixação, em especial ao Luciano, Lílian, Vanessa, Karen, Lys, Helisson, Giovani, Marcelo Muller, Marcelo Assumpção, Juliana Inaba, Adriana Invitti, Gustavo, Gus, Eliel, Fabiane, Luiza, André, André Luiz, Fernanda, Ana Claudia, Magda, Ju Osaki, Letícia, Stefania e Stefan pela amizade e incentivo.

Agradeço a todos os professores e colegas da Graduação em Ciências Biológicas que me acompanharam nestes quatro anos.

Ao quinteto fantástico Danillo, Lua, Reginaldo, Bia e Tânia pela amizade e por ter aceitado minha companhia nos elevadores da Reitoria.

As grandes amigas Anelis e Letícia por cada momento dentro e fora da Universidade, formamos um trio imbatível!

Gostaria muito de agradecer a minha grande amiga Manoela, por todo tempo, todos os conselhos, todas as festas, cada momento de alegria e de tristeza que dividiu comigo e que superamos juntas. Especialmente por todo o amor.

Agradeço a minha família, em especial à minha mãe, meu pai, a Cissa, Vitor, Artur e Rubens que sempre me deram amor, confiança, incentivo me ajudaram e torceram por mim sempre. Agradeço também a minha Vó Nelci que me acompanhou e também torceu por mim.

Finalmente gostaria de agradecer a minha Orientadora Profª Rose Adele Monteiro, que me proporcionou toda sua dedicação, carinho, confiança e paciência para que eu pudesse desenvolver este trabalho. Obrigada por cada dia de aprendizado e idéias, por todo incentivo, pela amizade, pelas risadas e cada segundo de convivência. Resumindo obrigada por todo amor e espero que ela conte sempre comigo assim como eu sempre contarei com ela. Muito obrigada.

SUMÁRIO

| | |
|---|------|
| AGRADECIMENTOS..... | iv |
| SUMÁRIO..... | vi |
| LISTA DE FIGURAS..... | viii |
| LISTA DE ABREVIATURAS..... | x |
| RESUMO..... | xi |
| 1. INTRODUÇÃO..... | 1 |
| 2. REVISÃO DE LITERATURA..... | 3 |
| 2.1 <i>Herbaspirillum seropedicae</i> | 3 |
| 2.2 FIXAÇÃO BIOLÓGICA DE NITROGÊNIO..... | 4 |
| 2.3 GENES <i>fix</i> | 5 |
| 2.4 REGULAÇÃO DA EXPRESSÃO DOS GENES <i>nif</i> E <i>fix</i> | 7 |
| 2.5 GENE <i>fixK</i> E SEUS HOMÓLOGOS..... | 7 |
| 2.6 FUNÇÃO DAS PROTEÍNAS FixK..... | 8 |
| 2.7 GENES <i>fixNOQP</i> | 9 |
| 2.7.1 Proteínas tipo citocromo c oxidases..... | 10 |
| 3. OBJETIVOS..... | 12 |
| 3.1 OBJETIVOS GERAIS..... | 12 |
| 3.2 OBJETIVOS ESPECÍFICOS..... | 12 |
| 4. MATERIAIS E MÉTODOS..... | 13 |
| 4.1 BACTÉRIAS E PLASMÍDEOS..... | 13 |
| 4.2 REAGENTES..... | 13 |
| 4.3 MEIOS DE CULTURA..... | 14 |
| 4.4 ANTIBIÓTICOS..... | 17 |
| 4.5 PURIFICAÇÃO DE PLASMÍDEOS..... | 17 |
| 4.6 SEQUENCIAMENTO DE DNA..... | 18 |
| 4.7 INSERÇÃO DO TRANSPOSON..... | 18 |
| 4.8 CLIVAGEM DO VETOR COM ENZIMAS DE RESTRIÇÃO PARA POSTERIOR LIGAÇÃO..... | 18 |
| 4.9 LIGAÇÃO DO FRAGMENTO DE DNA AOS VETORES DE EXPRESSÃO..... | 19 |
| 4.10. TRANSFORMAÇÃO BACTERIANA POR ELETROPORAÇÃO..... | 19 |
| 4.10.1 Preparo de células eletrocompetentes de <i>E. coli</i> e <i>H. seropedicae</i> | 19 |

| | |
|---|-----------|
| 4.10.2. Transformação bacteriana por eletroporação..... | 19 |
| 4.11 CARACTERIZAÇÃO DOS MUTANTES..... | 20 |
| 4.11.1 Hibridização de DNA genômico..... | 20 |
| 4.11.1.1 Preparo da sonda..... | 20 |
| 4.11.1.2 Transferência de DNA do gel para membrana de ‘nylon’..... | 20 |
| 4.11.1.3 Hibridização..... | 21 |
| 4.12. ATIVIDADE DA NITROGENASE..... | 21 |
| 4.13 DOSAGEM DE PROTEÍNAS..... | 22 |
| 4.14 ANÁLISE DAS SEQUÊNCIAS..... | 22 |
| 5. RESULTADOS..... | 24 |
| 5.1 ANÁLISE DAS SEQUÊNCIAS OBTIDAS..... | 24 |
| 5.2 INSERÇÃO DO TRANSPOSON..... | 37 |
| 5.3 SELEÇÃO DOS MUTANTES..... | 38 |
| 5.4 ANÁLISE E CONFIRMAÇÃO DOS MUTANTES..... | 40 |
| 5.5 DETERMINAÇÃO DO FENÓTIPO..... | 41 |
| 6. CONCLUSÃO..... | 43 |
| 7. REFERÊNCIAS BIBLIOGRÁFICAS..... | 44 |
| APÊNDICE 1..... | 51 |

LISTA DE FIGURAS

| | |
|---|----|
| FIGURA 1 - OPERON <i>fixISNOQPGH</i> de <i>H. seropedicae</i> | 10 |
| FIGURA 2 – CITOCROMO <i>cbb₃</i> OXIDASE..... | 11 |
| FIGURA 3a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA HIPOTÉTICA LOCALIZADA A MONTANTE DO GENE <i>fixI</i> | 25 |
| FIGURA 3b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA HIPOTÉTICA LOCALIZADA A MONTANTE DO GENE <i>fixI</i> | 25 |
| FIGURA 4a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixI</i> | 26 |
| FIGURA 4b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixI</i> | 27 |
| FIGURA 5a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixS</i> | 28 |
| FIGURA 5b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixS</i> | 28 |
| FIGURA 6a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixN</i> | 29 |
| FIGURA 6b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixN</i> | 29 |
| FIGURA 7a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixO</i> | 30 |
| FIGURA 7b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixO</i> | 31 |
| FIGURA 8a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixQ</i> | 31 |
| FIGURA 8b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixQ</i> | 32 |
| FIGURA 9a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FixP</i> | 32 |
| FIGURA 9b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA <i>FixP</i> | 33 |
| FIGURA 10a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DAS PROTEÍNAS <i>FixG</i> E <i>FixH</i> | 34 |
| FIGURA 10b - ALINHAMENTO DAS SEQUENCIAS DE AMINOÁCIDOS DEDUZIDAS DAS PROTEÍNAS <i>FixG</i> E <i>FixH</i> | 34 |
| FIGURA 11a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA <i>FNR</i> | 36 |

| | |
|--|----|
| FIGURA 11b - ALINHAMENTO DA SEQUENCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FNR | 36 |
| FIGURA 12 – ESQUEMA DE INSERÇÃO DO TRANSPOSON EZ:TN <TET-1> EM PLASMÍDEO..... | 38 |
| FIGURA 13 – ESQUEMA DE INSERÇÃO DO TRANSPOSON EZ:TN <TET-1> NO GENE <i>fixN</i> CROMOSSOMAL DE <i>H. seropedicae</i> SMR1..... | 39 |
| FIGURA 14 – LOCALIZAÇÃO DO TRANSPOSON EZ:TN <TET-1> NO GENE <i>fixN</i> E REGIÕES DE RESTRIÇÃO DA ENZIMA <i>Bgl</i> II..... | 40 |
| FIGURA 15 – PERFIL DE HIBRIDIZAÇÃO DAS ESTIRPES SELVAGEM SMR1 E DO MUTANTE <i>fixN</i> RAMM21..... | 41 |
| FIGURA 16 – ATIVIDADE DA NITROGENASE DAS ESTIRPES SELVAGEM SMR1 E MUTANTE RAMM21..... | 42 |

LISTA DE ABREVIATURAS

| | |
|-------------|--|
| A_{600nm} | - absorbância em 600 nm |
| Amp | - ampicilina |
| ARA | - atividade de redução do acetileno |
| ATP | - adenosina trifosfato |
| BSA | - albumina de soro bovino |
| Cm | - clorafenicol |
| dATP | - 5'-trifosfato de 2'- desoxiadenosina |
| dCTP | - 5'-trifosfato de 2'- desoxicitosina |
| dGTP | - 5'-trifosfato de 2'- desoxiguanosina |
| DMSO | - dimetilsulfóxido |
| DNA | - ácido desoxorribonucleico |
| D.O. | - densidade ótica |
| DTT | - ditionitrito |
| dTTP | - 5'-trifosfato de 2'- desoxitimidina |
| EDTA | - ácido etilenodiamino tetra acético |
| IPTG | - β -D-isopropil-tiogalactopiranosídeo |
| Kb | - quilopares de bases nucleotídicas |
| Km | - canamicina |
| Mg | - magnésio |
| N | - nitrogênio |
| N_2 | - dinitrogênio |
| NH_3 | - amônia |
| Nif | - fenótipo para fixação de nitrogênio |
| NO_2 | - dióxido de nitrogênio |
| p.b. | - pares de bases |
| rpm | - rotações por minuto |
| SDS | - dodecilsulfato de sódio |
| Sm | - estreptomicina |
| Tris | - tris (hidroximetil)-aminometano |
| UFC | - unidades formadoras de colônias |

RESUMO

Herbaspirillum seropedicae é um microrganismo diazotrófico, endofítico que se associa a diversas gramíneas de interesse econômico. O nitrogênio fixado por este organismo pode ser utilizado pela planta hospedeira, por isso é importante entender a fisiologia e genética do processo de fixação biológica de nitrogênio. A citocromo oxidase *cbb₃*, codificada pelos genes *fixNOQP*, é necessária para a respiração microaeróbica em vários organismos e em rizóbios, é essencial para a fixação biológica de nitrogênio. Estirpes mutantes nos genes *fixN* e *fixO* de *Bradyrhizobium japonicum* capazes de produzirem nódulos em soja, não são capazes de fixar nitrogênio, sugerindo que a citocromo *cbb₃* está envolvida com o transporte de elétrons neste organismo. O operon *fixNOQP* foi identificado pelo Projeto de Sequenciamento do Genoma do *H. seropedicae* (Genopar). A identificação da função desta enzima poderá esclarecer a via de transporte de elétrons neste organismo. O transposon EZ:TN <TET-1>, contendo um gene de resistência a tetraciclina, foi inserido no gene *fixN* do *H. seropedicae* da estirpe Smr1 por recombinação homóloga, produzindo a estirpe RAMM21. A atividade da nitrogenase deste mutante foi 40-50% da atividade da estirpe selvagem. Este resultado sugere que a citocromo oxidase *cbb₃* é importante para atividade de nitrogenase máxima em *H. seropedicae*.

1. INTRODUÇÃO

A atmosfera contém uma grande quantidade (cerca de 78% por volume) de nitrogênio molecular (N_2). Porém este grande reservatório de nitrogênio não está disponível para a maioria dos organismos vivos. A obtenção de nitrogênio da atmosfera requer a quebra de uma ligação tripla covalente de excepcional estabilidade, entre os dois átomos de nitrogênio para produzir amônia (NH_3) ou nitrato (NO_3^-). Tais reações, conhecidas como fixação do nitrogênio, podem ser obtidas por processo industrial e por processo natural (Taiz & Zeiger, 2004).

Sob elevadas temperaturas (cerca de $200^\circ C$) e alta pressão (cerca de 200 atmosferas), o N_2 combina-se com hidrogênio formando amônia. As condições extremas são necessárias para superar a alta energia de ativação da reação. Esta reação de fixação de nitrogênio, conhecida como processo *Haber-Bosh*, é ponto de partida para fabricação de muitos produtos para a indústria e agricultura. Indústrias em todo mundo produzem mais de 80×10^{12} g/ano de fertilizantes nitrogenados (Faostat, 2001).

A aplicação de grandes quantidades de nitrogênio pode aumentar a produção agrícola, entretanto, traz efeitos negativos, que incluem a poluição de águas subterrâneas e da atmosfera, pela liberação de NO_2^- e NO_3^- ; além dos altos custos de produção dos fertilizantes químicos decorrentes da queima de combustível fóssil na sua produção (Döbereiner, 1992).

Já os processos naturais fixam cerca de 190×10^{12} g ano⁻¹ de nitrogênio por meio de diferentes processos (Schlesinger, 1997). Descargas elétricas são responsáveis por aproximadamente 8% do nitrogênio fixado. Convertem o vapor de água e oxigênio em radicais hidroxilas livres altamente reativos, em átomos de hidrogênio livre e em átomos de oxigênio livre, que atacam o nitrogênio molecular (N_2) para formar ácido nítrico (HNO_3). Posteriormente, este ácido nítrico precipita-se sobre a Terra junto com a chuva (Taiz & Zeiger, 2004).

Outros 2% deste nitrogênio fixado é derivado de reações fotoquímicas entre o óxido nítrico gasoso (NO) e o ozônio (O_3), produzindo ácido nítrico (HNO_3).

Noventa por cento do nitrogênio naturalmente fixado resultam da fixação biológica, nos quais certos organismos procariotos e cianobactérias fixam o N_2 em amônio (NH_3).

A descoberta de várias espécies de bactérias fixadoras de nitrogênio, diazotróficas, em associações naturais com raízes de gramíneas (Dobereiner, 1988), intensificou o interesse na fixação biológica de nitrogênio.

Entre os organismos diazotróficos e endofíticos estudados estão aqueles do gênero *Herbaspirillum*. Estudos sobre sua ecologia têm mostrado que esta bactéria pode ser isolada do interior de raízes, hastes e folhas de muitas gramíneas (Baldani et al., 1992; Dobereiner et al., 1992; Olivares et al., 1996).

Vários estudos mostram que a interação *Herbaspirillum*-planta pode trazer mútuos benefícios. Baldani e colaboradores (1995) inocularam diferentes estirpes de *Herbaspirillum* em sementes de arroz e observaram que a estirpe Z94 contribuiu com até 54% do aumento do conteúdo de nitrogênio total acumulado pela planta; outras estirpes também fixaram aproximadamente 30% de nitrogênio a mais. A história de certas regiões produtoras de cana-de-açúcar no Brasil mostra áreas cultivadas por mais de cem anos sem adubação nitrogenada e sugere que a fixação biológica de nitrogênio, principalmente pelos endofíticos *Herbaspirillum* ssp e *Gluconacetobacter diazotrophicus*, possa ter contribuído significativamente para o metabolismo nitrogenado da planta (Döbereiner, 1992). Estes resultados indicam que o *H. serapedicae* tem grande potencial como biofertilizante nitrogenado, além de estimular o desenvolvimento das plantas pela produção de fitohormônios (Olivares et al., 1997).

O processo de colonização pelo *H. serapedicae* e as vias envolvidas na fixação de nitrogênio ainda não foram totalmente esclarecidos. Vários genes podem estar envolvidos nesse processo, como por exemplo, os genes *nif*. O complexo enzimático da nitrogenase é responsável pela conversão do N_2 a $2NH_3$. Esta reação é exergônica e estimulada em condições limitantes de oxigênio. Assim como a proteína NifA, que é ativadora dos genes *nif*, a transcrição de outras proteínas podem ser induzidas pela ausência de oxigênio ou de amônia, para que o organismo suporte a respiração em condições limitantes de oxigênio. A Citocromo *cbb₃* foi descrita como uma proteína transportadora de elétrons em vários organismos formadores de nódulos, e é codificada pelos genes *fixNOQP*.

O *H. serapedicae* possui estes genes e podem estar envolvidos na respiração sob condições de fixação de nitrogênio.

Neste trabalho observamos que o mutante *fixN* apresentou atividade da nitrogenase reduzida em relação a estirpe selvagem.

2. REVISÃO DE LITERATURA

2.1 *Herbaspirillum seropedicae*

Herbaspirillum seropedicae é uma bactéria diazotrófica endofítica, gram-negativa, vibrióide, membro da subdivisão β das Proteobactérias (Baldani et al., 1986). Essa bactéria foi encontrada associada ao milho, arroz, sorgo e cana de açúcar (Baldani et al., 1984; Boddey et al., 1995).

Estudos realizados por OLIVARES e colaboradores (1995) concluíram que a infecção por *Herbaspirillum ssp.* ocorre através da ligação da bactéria à superfície da planta, seguido de proliferação preferencialmente nas raízes secundárias e ferimentos, penetração e espalhamento da bactéria através dos espaços intercelulares e feixes vasculares das partes aéreas com subsequente colonização e estabelecimento nos vasos do xilema.

A bioquímica e biologia molecular da fixação de nitrogênio em *H. seropedicae* têm sido intensamente estudadas, isso levou a identificação de diversos genes envolvidos no processo. Dentre os genes isolados e caracterizados está o *nifA* que codifica para a proteína NifA (Souza et al., 1991). Anteriormente acreditava-se que os genes *nif* neste organismo estavam organizados em duas diferentes regiões, na região I encontravam-se os genes *nifA* e *nifB* (Souza et al., 1991a), e na região II o operon *nifHDKENXofr1 orf2 orf3* (Klassen et al., 1999). Posteriormente, a sequência completa do *H. seropedicae* (GENOPAR) revelou que estas duas regiões são contíguas ocupando cerca de 40 Kb.

A regulação da expressão do gene *nifA* envolve mecanismos variáveis, que depende da fisiologia de cada diazotrofo. Nos organismos de vida livre o nível de nitrogênio associado é o fator mais importante da regulação, enquanto que nos organismos simbióticos o fator primordial é o oxigênio (Merrick, 1993).

A transcrição do gene *nifA* de *H. seropedicae* depende do fator σ^N e é ativada pela proteína NifA. Na região promotora desse gene estão presentes os sítios de ligação para a proteína NifA e para a proteína NtrC. Foram ainda identificados uma sequência semelhante à sequência consenso "anaerobox" localizada após o início de tradução e um sítio de ligação para a proteína IHF (Souza et al, 1991b). A transcrição do gene *nifA* de *H. seropedicae* é ativada pela proteína NtrC, sendo assim regulada por amônia e não por oxigênio (Souza et al., 1995).

A proteína NifA de *H. seropedicae* possui aproximadamente 61 kDa e apresenta três domínios. Sua região N-terminal mostra baixa similaridade nas regiões C-terminal e central (Monteiro et al., 1999). Como NifA de *H. seropedicae* é inativada por altas concentrações de oxigênio, esta capacidade pode ser devido a presença de um interdomínio entre as regiões central e C-terminal, possivelmente envolvido com sensibilidade a oxigênio em outras proteínas NifA (Fischer et al., 1988). Resultados posteriores confirmaram que os níveis de oxigênio controlam a atividade de NifA. O domínio N-terminal está envolvido no controle da atividade da própria NifA por amônio, uma vez que a proteína sem a região N-terminal continua ativa mesmo na presença de altas concentrações de amônio (Souza et al., 1999). A proteína NifA de *H. seropedicae* é inativa quando expressa em *E. coli*, independente dos níveis de oxigênio e/ou amônia. A proteína N-truncada, entretanto é ativa (Monteiro et al., 1999). A proteína PII participa da modulação da atividade da NifA de *H. seropedicae* (Souza et al., 1999).

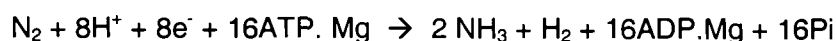
Para investigar a possibilidade da proteína NifA de *H. seropedicae* receber um sinal de uma proteína sensora de oxigênio, semelhante ao sistema descrito para *K. pneumoniae*, Monteiro e colaboradores (2003) realizaram ensaios *in vivo* para avaliar a atividade da proteína NifA expressa em células de *E. coli* mutantes para o gene *fnr*. Os resultados obtidos com a estirpe mutante *fnr* mostraram que a proteína de fusão His-tag-central+C-terminal é inativa, independente da presença ou ausência de oxigênio. Estes resultados sugerem uma relação funcional entre a proteína NifA de *H. seropedicae* e a proteína FNR (Monteiro et al., 2001).

2.2 FIXAÇÃO BIOLÓGICA DE NITROGÊNIO

Os organismos diazotróficos são capazes de produzir amônia a partir do nitrogênio molecular. Todos os microrganismos não diazotrofos dependem direta ou indiretamente dos diazotróficos para suplementação de compostos de nitrogênio, que são utilizados para síntese de ácidos nucleicos e proteínas, através de vias bioquímicas. Anualmente, os microrganismos diazotróficos adicionam cerca de 65% de nitrogênio à Terra, enquanto que o nitrogênio industrialmente fixado contribui com 25% do volume (Newton, 2000).

A fixação biológica de nitrogênio requer condições favoráveis tais como baixas concentrações de nitrogênio fixado, baixas tensões de oxigênio e temperatura adequada (Postgate, 1982).

A redução do dinitrogênio à amônia requer a transferência de seis elétrons ao N_2 e à redução de oito prótons formando também H_2 . O complexo da nitrogenase catalisa esta reação (Halbleib & Ludden, 2000). A estequiometria é mostrada abaixo (Simpson & Burris, 1984):



O complexo da nitrogenase possui dois componentes, a proteína Fe e a proteína MoFe. A proteína Fe, produto do gene *nifH*, reduz a proteína MoFe e é irreversivelmente inativada por O_2 .

A proteína MoFe, produto dos genes *nifD* e *nifK*, apresenta quatro subunidades (α_2 e β_2) e dois grupos Mo-Fe-S. Ela também é inativada pelo oxigênio (Burgess & Lowe, 1996). Quando a quantidade de nitrogênio fixada pela planta é suficientemente alta, a expressão destes genes é reprimida. Em *Klebsiella pneumoniae* pelo menos 20 genes *nif* são necessários para a fixação de nitrogênio (Merrick, 1992).

Como a nitrogenase requer alta quantidade de elétrons, vias específicas de transporte de elétrons estão presentes nos diversos diazotrofos dependendo das condições ambientais fornecidas. Em baixas concentrações de O_2 diferentes vias devem ser acionadas para que o complexo da nitrogenase continue ativo. O sistema respiratório de muitas bactérias possui diferentes proteínas oxidases terminais, possibilitando que a bactéria se desenvolva em diferentes tensões de oxigênio (Zufferey et al., 1996).

2.3 GENES *fix*

Os genes de fixação de nitrogênio de bactérias simbiontes podem ser basicamente divididos em três grupos: genes *nod*, *nif* e *fix*. Em rizóbios os produtos dos genes *nod* são necessários nos passos de formação do nódulo. Os genes *nif* de rizóbios são estruturalmente homólogos aos 20 genes *nif* de *K. pneumoniae* e estão diretamente ligados às proteínas que convertem o nitrogênio atmosférico a amônia. Os genes *fix* são essenciais para a fixação do nitrogênio, mas não têm homologia com os genes de fixação

de nitrogênio de *K. pneumoniae*, que são homólogos aos genes *nif*. Esta definição geral traz três consequências. A primeira, os genes *fix* representam uma classe muito heterogênea incluindo genes envolvidos no desenvolvimento e metabolismo das bactérias. Segundo, genes originalmente nomeados *fix* podem ser renomeados quando sua função for melhor conhecida. Terceiro, genes inicialmente identificados no contexto da fixação de nitrogênio simbiótica podem participar de uma série de processos não relacionados com fixação, tornando este nome inapropriado (Fisher, 1994).

Os genes *fix* e *nif* de *Synorhizobium meliloti*, *B. japonicum* e *A. caulinodans* são organizados em diferentes grupos cuja estrutura e localização genômica são espécies-específicos. Em *R. meliloti* estes dois grupos são nomeados cluster I (*nifHDKE*, *nifN*, *fixABCX* *nifA* *nifB* *frdX*) e cluster II (*fixLJ*, *fixK*, *fixNOQP*, *fixGHIS*) (Fischer, 1994).

Mutações em qualquer gene do operon *fixABCX* de *S. meliloti*, *B. japonicum* e *A. caulinodans* aboliu completamente a capacidade de fixação do mutante, isto sugeriu que eles estão diretamente ligados com o transporte de elétrons para a nitrogenase (Earl et al., 1987).

A análise da sequência de bases dos genes *fixGHIS* sugeriu que os produtos destes genes são provavelmente proteínas transmembranas. A proteína FixG pode estar envolvida no processo redox, pois contém dois conjuntos de cisteínas típicos, de proteínas com centro ferro-enxofre. A proteína FixI é homóloga à subunidade catalítica de ATPases de bactérias e eucariotos. A função global do operon e seu produto na fixação de nitrogênio ainda não estão completamente esclarecidos (Kahn et al., 1989).

Os genes *fixLJ* de *S. meliloti* foram originalmente identificados como geneticamente responsáveis pela ativação dos operons *fixNOQP* e *fixGHIS* assim como do gene *nifA* em condições limitantes de oxigênio. Deleções nos genes *fixL* e *fixJ* resultaram em nódulos simbioticamente ineficientes (Fix⁻) em *S. meliloti* e *A. caulinodans*, e com atividade da nitrogenase drasticamente diminuída (David et al., 1988). Similarmente, estirpes mutantes de *B. japonicum* nestes genes mostraram uma redução de 90% na fixação simbiótica de nitrogênio em relação à estirpe selvagem, além de serem incapazes de crescerem anaeróbicamente indicando que o nitrato não é o receptor final de elétrons nestes mutantes (Anthamatten et al., 1992).

2.4 REGULAÇÃO DA EXPRESSÃO DOS GENES *nif* E *fix*

A fixação de nitrogênio requer uma demanda energética muito alta para a redução de dinitrogênio a amônia (o complexo enzimático da nitrogenase utiliza 16 mol de ATP para reduzir 1 mol de N_2 [Merrick, 1992]). Além disso, há uma grande sensibilidade da nitrogenase ao oxigênio. Estes fatores explicam o fato dos diazotrofos estudados terem desenvolvido sofisticados mecanismos de regulação que garantam um alto controle da síntese e atividade da nitrogenase, para evitar gastos energéticos desnecessários. Devido a este controle, a expressão dos genes da fixação de nitrogênio é regulada por cascatas de genes. A sua ação em conjunto permite que a bactéria diferencie condições ambientais ótimas para fixação de nitrogênio e transmita esta informação ao nível de expressão genética (Fischer, 1994).

Os níveis ambientais de oxigênio e de nitrogênio são sinais críticos para a regulação da expressão dos genes *nif* em *K. pneumoniae*. Este controle fisiológico duplo é mediado via um sistema regulatório geral de nitrogênio (*ntr*) e genes regulatórios *nif*-específicos, *nifLA* (Merrick, 1992).

A proteína FixK em rizóbio está envolvida na regulação da expressão de vários genes necessários para respiração no nódulo, ambiente de baixos níveis de oxigênio. E, assim como os seus homólogos *fnr* e *crp* de *E. coli*, o gene *fixK* de *R. meliloti* regula a sua própria expressão negativamente (Foussard et al., 1997). *FixK* de *R. meliloti* é parte de uma cascata regulatória onde a proteína FixL atua como um sensor de baixas concentrações de oxigênio, se auto fosforila, e depois transfere o grupo fosfato ao regulador FixJ, que finalmente ativa a expressão de *fixK* (Batut & Boistart, 1994).

2.5 GENE *fixK* E SEUS HOMÓLOGOS

Em *S. meliloti* o gene *fixK* foi identificado entre os operons *fixLJ* e *fixNOQP*. Um gene tipo *fixK* (*fixK₂*) foi identificado a jusante do operon *fixLJ* em *B. japonicum* (Fischer, 1994). Outro gene tipo *fixK* já havia sido descrito anteriormente em *B. japonicum*, apesar de que nenhuma função lhe tenha sido atribuída (Anthamatten et al., 1992). Para diferenciar do gene *fixK₂*, que é controlado por *fixJ*, este gene foi identificado como *fixK₁* de *B. japonicum*. Em *A. caulinodans*, o gene tipo *fixK* está presente a montante dos genes *fixLJ*, mas com orientação de transcrição oposta (Kaminski et al., 1991). Existe ainda

outro gene tipo *fixK* em *A. caulinodans*, mas com função diferente daquela do gene *fixK* encontrado a montante de *fixLJ*. Em *Rhizobium leguminosarum* bv. *viciae* um gene tipo *fixK* (*fnrN*) foi identificado com base na capacidade de seu produto de ativar a expressão de uma fusão heteróloga *fixN::lacZ* sob condições de microaerobiose e de complementar parcialmente um mutante *fixJ* de *S. meliloti* (Collona-Romano et al., 1990).

Mutações do gene *fixK* de *S. meliloti* e *A. caulinodans* e *fixK₂* de *B. japonicum* eliminam completamente a fixação de nitrogênio nestes organismos. Mutantes de *fixK₂* de *B. japonicum* são deficientes na respiração por nitrato (Anthamatten et al., 1991). Não existem dados disponíveis para respiração por nitrato em *A. caulinodans*. Mutantes *fixK₁* em *B. japonicum* e *fnrN* em *R. leguminosarum* bv. *viciae* ainda são capazes de fixar nitrogênio, mas a atividade de nitrogenase é menor no último caso (Fischer, 1994).

Além de serem regulados positivamente, os genes *fixK* de *S. meliloti* e *fixK₂* de *B. japonicum* estão sujeitos a uma auto-regulação negativa (Batut et al., 1989). Em 1997, Foussard e colaboradores demonstraram que a auto-regulação negativa de *fixK* de *S. meliloti* não ocorre de forma direta, mas envolve o produto de um novo gene identificado recentemente, *fixT*, cuja expressão depende de *fixK*. A inativação de *fixT* resultou na ativação da expressão de *fixK* em condições de microaerobiose.

2.6 FUNÇÃO DAS PROTEÍNAS FixK

A similaridade estrutural entre *FixK₁* de *B. japonicum* e *FnrN* de *R. leguminosarum* bv. *viciae* com *FNR* de *E. coli* é também refletida a nível funcional. Tanto *FixK₁* quanto *FnrN* podem substituir *FNR* de *E. coli* com relação a ativação de promotores *FNR*-dependentes, e esta ativação é dependente da presença de condições microaeróbicas e anaeróbicas (Anthamatten et al., 1992). *FNR* pode ativar uma fusão *fixN::lacZ* de *S. meliloti* em *E. coli* e em *R. leguminosarum* bv. *viciae*. Em testes de complementação homóloga foi mostrado que *fixK₁* de *B. japonicum* é capaz de substituir *fixK₂* na regulação da respiração anaeróbica por nitrato e fixação simbiótica de nitrogênio (Anthamatten et al., 1992).

Uma função comum das proteínas do tipo *FixK* de *S. meliloti*, *A. caulinodans* e *B. japonicum* é seu controle da expressão do operon *fixNOQP* (Batut et al., 1989; Mandon et al., 1994). Possivelmente, as proteínas *FixK* de *S. meliloti* e *FixK₂* de *B. japonicum* também controlam a expressão do operon *fixGHIS*, que está situado imediatamente a

jusante do operon *fixNOQP* em ambos os organismos. Esta hipótese é reforçada pela presença de um potencial sítio de ligação para FixK a montante de *fixG* (Preisig et al. 1993).

O número crescente de prováveis genes e operons regulados por FixK permitiu determinar a partir das diferentes regiões promotoras um potencial sítio de ligação de FixK (Fisher, 1994). Há pouca informação disponível com relação a promotores de rizóbios que não sejam dependentes de σ^{54} . Uma sequência conservada (5' **-TTGA-C-GATCAA-G-3'**) foi localizada a montante dos genes controlados positivamente por FixK. Esta sequência é bastante similar à sequência do sítio de ligação para FNR de *E. coli* ("*FNR box*": 5' **-AAA-TTGAT---ATCAA-TTT-3'**) (Eiglmeier et al., 1989). Esta similaridade não é surpreendente, já que há uma alta similaridade entre o provável motivo de ligação do DNA de FNR e das proteínas FixK (Fischer, 1994). Em *H. seropedicae* dois prováveis promotores foram identificados a montante do provável gene *fnr*, um deles dependente do fator σ^{70} (promotor de expressão constitutiva) e outro dependente do fator σ^N . Neste mesmo trabalho foi identificada uma provável região de ligação de FNR – FNR Box, sugerindo uma auto-regulação pela proteína FNR (Raeder, 2003).

2.7 GENES *fixNOQP*

O operon *fixNOQP* de bactérias que são capazes de se associar com plantas (por exemplo aquelas dos gêneros *Rhizobium*, *Azorhizobium* e *Bradyrhizobium*), codifica para subunidades da citocromo oxidase *cbb₃* (Kahn et al, 1993). A citocromo *cbb₃* é uma oxidase respiratória, que possui três subunidades. A subunidade I (FixN), possui um grupo heme e um átomo de cobre ligados. Um citocromo c monoheme (FixO) é a subunidade II; a subunidade III (FixP) é um citocromo C diheme, (Garcia-Horsman et al., 1994). A citocromo oxidase do tipo *cbb₃* suporta a respiração de organismos como *B. japonicum* em condições limitantes de oxigênio e é necessária para fixação biológica de nitrogênio quando em simbiose (Zufferey et al., 1996).

Em *H. seropedicae* os genes *fixNOQP* foram identificados a jusante aos genes *fixI* e *fixS*, contendo 3176 pb e a montante do gene *fixG* (figura1).

FIGURA 1 - OPERON *fixLSNOQPGH* de *H. seropedicae*



FIGURA 1 – Organização estrutural do operon *fixLSNOQPGH* de *H. seropedicae*.

Os genes *fixNOQP* foram primeiramente identificados em *Synorhizobium meliloti* ligados aos genes regulatórios *fixLJ* e *fixK*, os quais induzem a expressão de *fixNOQP* em anaerobiose assim como a do gene *nifA*. Genes homólogos foram identificados em *B. japonicum*, *A. caulinodans*, e *Rhizobium leguminosarum* bv. *viciae*. (Fischer, 1994).

Mutantes *fixN* de *Azorhizobium caulinodans* tiveram a atividade da nitrogenase reduzida em 40 a 50% (Mandon et al., 1994).

Mutantes *B. japonicum* no gene *fixN* ou *fixO* não são capazes de fixar nitrogênio em condições baixas de oxigênio ou em anaerobiose (Lopez et al., 2001).

2.7.1 Proteínas tipo citocromo C oxidases

O sistema respiratório das bactérias é usualmente munido de várias oxidases terminais com diferentes afinidades por O₂, fazendo com que a bactéria possa respirar em várias tensões de oxigênio. A *cbb₃* oxidase é uma citocromo oxidase que contém três subunidades: uma membrana-integrada b-oxidase (subunidade I) com um centro de ferro de baixo spin e outro centro binuclear de alto spin Fe-Cu, e duas subunidades membrana-ancoradas contendo citocromos do tipo c (figura 2). Em *B. japonicum*, os genes *fixNOQP* expressos geralmente em condições limitantes de oxigênio codificam para a citocromo *cbb₃* oxidase, são genes essenciais para a fixação de nitrogênio durante a simbiose (Zufferey, 1996). Mutantes de *B. Japonicum* no gene *fixN* ou *fixO*, diferentemente do mutante *fixQ*, não apresentaram a citocromo *cbb₃* em sua membrana, indicando quais os genes essenciais para a formação desta proteína neste organismo. Nos mutantes *fixP*,

somente a região FixP da proteína foi deletada e houve produção normal das subunidades codificadas pelos genes *fixN* e *fixO* (Zufferey, 1996).

FIGURA 2 – CITOCROMO *cbb₃* OXIDASE

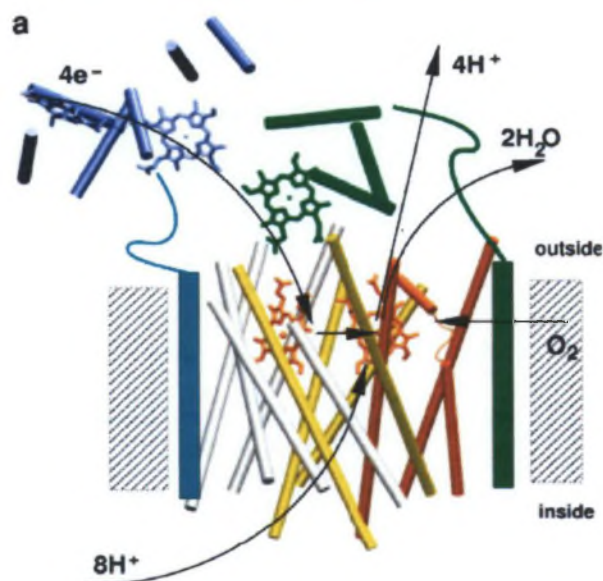


FIGURA 2 - Vista total do modelo da oxidase *cbb₃* de *B. japonicum*, a parte hachurada é a membrana em vista lateral. O citoplasma está para baixo, e o espaço periplasmático está acima. A subunidade catalítica é mostrada ao meio, com as hélices coloridas de acordo com o grau de conservação: laranja para as hélices mais conservadas, branca para menos conservadas, e o amarelo para um grau intermediário de similaridade da sequência. O domínio solúvel da subunidade O (verde) e da subunidade P (azul-claro) estão representados com suas hélices transmembrana. A região mais conservada representa a subunidade N. A subunidade Q é representada por uma hélice transmembrana, que aparece em marrom. As flechas então indicando o movimento de eletros através da membrana (Sharma et al; 2006).

3. OBJETIVOS

3.1 OBJETIVOS GERAIS

O *H. seropedicae* é um organismo fixador de nitrogênio e tem potencial como biofertilizante nitrogenado. A determinação da função do operon *fixNOQP* na fixação de N₂ neste organismo ampliará o entendimento do mecanismo de controle da fixação de nitrogênio. Para tanto, a estratégia utilizada foi a mutação do gene *fixN* do *H. seropedicae* através da inserção de um transposon, e análise fenotípica do mutante obtido.

3.2 OBJETIVOS ESPECÍFICOS

- Construção do plasmídeo contendo o gene *fixN* mutagenizado;
- Obtenção dos mutantes *fixN* de *H. seropedicae*;
- Determinar atividade da nitrogenase do mutante *fixN* de *H. seropedicae*.

4.MATERIAL E MÉTODOS

4.1 BACTÉRIAS E PLASMÍDEOS

As estirpes de bactérias e plasmídeos que foram utilizados neste trabalho estão relacionadas na Tabela 1.

TABELA 1 – ESTIRPES DE BACTÉRIAS E PLASMÍDEOS UTILIZADOS

| Estirpe ou plasmídeo | Genótipo ou Características relevantes | Referência |
|-----------------------|--|------------------------------|
| <i>E. coli</i> | | |
| TOP10 | F- <i>mcrA</i> $\Delta(mrr-hsdRMS-crBC)$ $\phi80lacZ\Delta$ M15 $\Delta lacX74$ <i>recA1</i> <i>ara</i> Δ 139 $\Delta(ara-leu)$ 7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str ^R) <i>endA1</i> <i>nupG</i> | INVITROGEN |
| <i>H. seropedicae</i> | | |
| Smr1 | Sm ^r , NIF ⁺ (derivada da estirpe Z78) | PEDROSA <i>et al.</i> , 1997 |
| Plasmídeos | | |
| HS17058H11 | Amp ^R | Projeto Genopar |
| HS17058H11Tc | Tc ^R Amp ^R | Este Trabalho |
| HS17058H11TcCm | Cm ^R Tc ^R Amp ^R | Este Trabalho |

4.2 REAGENTES

Os antibióticos, trizma-base, o-nitrofenil β -D-galactopiranosídeo (ONPG) e ficoll foram adquiridos da Sigma Chemical Company. As enzimas modificadoras de DNA e RNase foram adquiridas de GE Healtecure, Boehringer Mannheim Biochemicals ou INVITROGEN. Para a análise de DNA foi utilizado o padrão de massa molecular “O’GeneRuler^TM 1kb DNA Ladder” da Fermentas. Fenol, etanol, clorofórmio, isopropanol

e álcool isoamílico foram adquiridos de Merck. Extrato de levedura, triptona, ágar bacteriológico foram adquiridos da Biobrás ou INVITROGEN. Os demais reagentes foram para análise foram de alto grau de pureza.

4.3 MEIOS DE CULTURA

Os meios de cultura utilizados no cultivo das estirpes de *E. coli* foram o Luria-Broth (LB e LA), SOB e SOC.

O meio LB apresenta a seguinte composição:

| | Gramas / Litro |
|---------------------|----------------|
| Extrato de levedura | 5,0 |
| Cloreto de sódio | 10,0 |
| Triptona | 10,0 |

O meio sólido foi obtido pela adição de ágar na concentração de 15 g/L ao meio líquido, sendo denominado LA.

O meio SOB apresenta a composição descrita na tabela abaixo:

| | Gramas / Litro |
|---------------------|----------------|
| Bacto triptona | 20,0 |
| Extrato de Levedura | 5,0 |
| Cloreto de sódio | 0,584 |
| Cloreto de potássio | 0,186 |

Meio SOC:

| | Gramas / Litro |
|---------------------|----------------|
| Triptona | 20 |
| Extrato de Levedura | 5 |
| Cloreto de sódio | 0,5 |
| Cloreto de potássio | 0,186 |
| MgCl ₂ | 0,94 |
| MgSO ₄ | 1,2 |
| Glucose | 3,6 |

Todos os meios foram autoclavados e o pH acertado para 7,0 com NaOH.

As estirpes de *H. seropedicae* foram cultivadas a 30°C, sob agitação a 120 rpm, durante 16 horas em meio NFbHP (KLASSEN *et al.*, 1997), utilizando estreptomicina como antibiótico selecionador. O meio NFbHP possui a seguinte composição:

| | |
|--|-----------------------------|
| MgSO ₄ .7H ₂ O | 2 x 10 ⁻¹ g/L |
| NaCl | 1 x 10 ⁻¹ g/L |
| CaCl ₂ | 2 x 10 ⁻² g/L |
| Ácido nitrilo-triacético | 5,6 x 10 ⁻² g/L |
| FeSO ₄ . 7H ₂ O | 2 x 10 ⁻² g/L |
| Malato de sódio | 5 g/L |
| Biotina | 1 x 10 ⁻⁴ g/L |
| Na ₂ MoO ₄ .2 H ₂ O | 2 x 10 ⁻³ g/L |
| MnSO ₄ . H ₂ O | 2,35 x 10 ⁻³ g/L |
| H ₃ BO ₃ | 2,8 x 10 ⁻³ g/L |
| CuSO ₄ .5 H ₂ O | 8 x 10 ⁻⁵ g/L |
| ZnSO ₄ .7 H ₂ O | 2,4 x 10 ⁻⁴ g/L |

No momento do uso foi adicionado ao meio NFbHP, 50mmol/L de solução estéril de fosfatos (159,5g/L de KH₂PO₄ e 17,8g/L de K₂HPO₄) e 20mmol/L de cloreto de amônio (NFbHPN) ou 5mmol/L de glutamato de sódio (NFbHPG), também estéreis.

4.4 ANTIBIÓTICOS

Os antibióticos utilizados foram:

| | Solução estoque (mg/mL) | Concent. final (µg/mL) |
|--------------------|-------------------------|------------------------|
| Ampicilina (Amp) | 100 | 200 |
| Cloranfenicol (Cm) | 30 | 120 |
| Tetraciclina (Tc) | 10 | 10 |
| Estreptomicina | 80 | 80 |

As soluções estoques de antibióticos, exceto cloranfenicol e tetraciclina, foram preparadas em água destilada autoclavada e esterelizadas por filtração (Millipore HAWP 0,23µm). A solução estoque de cloranfenicol foi preparada em água e tetraciclina foi preparada em etanol 70%. Todas as soluções de antibióticos foram mantidas a -20°C.

4.5 PURIFICAÇÃO DE PLASMÍDEOS

O isolamento dos plasmídeos de estirpe TOP10 de *E.coli* foi feito pelo método de lise alcalina (SAMBROOK *et al.*, 1989) com modificações. Três mililitros de uma cultura incubada durante aproximadamente 12 horas foram centrifugadas a 13000 rpm por 2 minutos. O sedimento de células foi ressuscitado em 200µL de tampão GET (50mmol/L glucose, 25mmol/L Tris-HCl pH 8,0, 10mmol/L de EDTA pH8,0). A lise foi então efetuada com 200µL de solução de SDS 1% e NaOH 0,2mol/L adicionados vagarosamente. Em seguida 200µL de acetato de potássio 3mol/L pH 4,8 foram acrescentados e após homogeneização a mistura foi mantida durante 10 minutos no gelo. Após esse período, as amostras foram centrifugadas a 13000 rpm por 5 minutos. O sobrenadante contendo o DNA plasmidial foi separado e precipitado pela adição de 600µL de isopropanol. Depois as amostras foram novamente centrifugadas a 13000 rpm por 5 minutos. Após centrifugação o precipitado foi lavado com etanol 80%, e seco a vácuo. O precipitado foi dissolvido em água.

4.6 SEQUENCIAMENTO DE DNA

Ao DNA dupla fita purificado (0,5-1,0 μ g) foram adicionados 3 μ L de mistura para seqüenciamento ET mix (Amersham Biosciences) e 1 pmol de oligonucleotídeo reverso ou universal num volume final de 7,5 μ L. Esta solução então foi submetida a uma reação de amplificação do DNA para seqüenciamento utilizando termociclador Applied Biosystem 9700 com a seguinte programação (30 ciclos):

Etapa 1: 96°C 20seg

Etapa 2: 60°C 1min 30seg

O produto da reação de amplificação de DNA para seqüenciamento foi transferido para um tubo do tipo Eppendorf de 500 μ L e foram adicionados 12,5 μ L de H₂O, 2 μ L de acetato de amônio 7,5 mol/L e 60 μ L de etanol absoluto. O sistema foi centrifugado (13000 rpm, 30 minutos), o sobrenadante foi descartado e o precipitado lavado com 450 μ L de etanol 80%. Após centrifugação o sobrenadante foi novamente descartado e o material seco a vácuo. A amostra foi desnaturada por 2 minutos a uma temperatura de 96°C e submetida à eletroforese em seqüenciador automático ABI377 (Applied Biosystems).

4.7 INSERÇÃO DO TRANSPOSON

A reação de inserção de transposon, contendo um gene que confere resistência à tetraciclina, foi feita *in vitro*. O protocolo utilizado é recomendado pelo fabricante (EPICENTRE).

4.8 CLIVAGEM DO VETOR COM ENZIMAS DE RESTRIÇÃO PARA POSTERIOR LIGAÇÃO

A digestão do vetor foi feita segundo protocolo descrito por Sambrook e colaboradores (1989) ou recomendação do fabricante da enzima de restrição.

4.9 LIGAÇÃO DO FRAGMENTO DE DNA AOS VETORES DE EXPRESSÃO

O fragmento de DNA de interesse e o vetor tratado com as enzimas de restrição apropriadas foram ligados baseado em um sistema. A ligação de fragmento de DNA e vetor foi feita na presença de tampão T4 DNA ligase 1X (50mmol/L Tris-HCl pH 7,6, 10mmol/L MgCl₂, 1mmol/L DTT, 0,5mmol/L ATP e 5%PEG 8000) e 1U de T4 DNA ligase (Sambrook et al., 1989). Foi utilizada uma relação molar 10:1 de inserto e vetor. O sistema foi incubado a aproximadamente 8°C durante 12 horas, e então utilizado para a transformação bacteriana.

4.10. TRANSFORMAÇÃO BACTERIANA POR ELETROPORAÇÃO

4.10.1 Preparo de células eletrocompetentes de *E. coli* e *H. seropedicae*

Quarenta mL de meio líquido foram inoculados com 400μL de uma cultura (D.O. 1,0) de *E. coli* (meio SOB) ou *H. seropedicae* (meio NFbHPN) crescida durante 12h. A cultura foi incubada sob agitação a 37°C para *E.coli* e 30°C para *H. seropedicae* até atingir uma D.O₆₀₀ entre 0,5 e 1,0. Após crescimento, a cultura foi mantida no gelo durante 30 minutos e, em seguida, centrifugada a 4000 rpm por 5 minutos. As células foram lavadas duas vezes com 40mL de H₂O estéril gelada (4°C), e então ressuspensas em 10 mL de glicerol 15% gelado. Após centrifugação a 4000 rpm por 5 minutos, o sedimento de células foi ressuspenso em 200μL de glicerol 5%. A suspensão celular foi fracionada em alíquotas de 40μL que foram armazenadas a -70°C.

4.10.2. Transformação bacteriana por eletroporação

O DNA foi adicionado a 40μL da suspensão de células eletrocompetentes, e incubado sobre gelo por 1 minuto. Em seguida essas células foram transferidas para uma cubeta de eletroporação e submetidas a um campo elétrico (4KV/cm, 330μF) para permitir a entrada do plasmídeo na célula bacteriana. Após a eletroporação, as células de *H. seropedicae* foram ressuspensas em 1mL de meio NFbHPN e incubadas a 30°C, sob agitação por 1 hora. Para *E. coli* o meio usado foi o SOC e a incubação feita a 37°C. Em

seguida, alíquotas da suspensão foram plaqueadas em meio NFbHPN sólido, para *H. seropedicae*, e LA, para *E. coli*, contendo os antibióticos adequados à seleção dos transformantes de interesse.

4.11 CARACTERIZAÇÃO DOS MUTANTES

4.11.1 Hibridização de DNA genômico

4.11.1.1 Preparo da sonda

O DNA do plasmídeo HS17058H11 digerido com a enzima *Bgl*II foi utilizado como sonda nos ensaios de hibridização. Cerca de 25 ng de DNA-sonda foram submetidos a desnaturação pelo calor em banho fervente por 10 minutos e rapidamente resfriado em banho de gelo. O material desnaturado foi incubado por 30 minutos a 37°C com uma mistura contendo nucleotídeos (dATP, dGTP, dTTP) na concentração final de 25 µmol/L de cada nucleotídeo, 10 µCi de [³²P-α] dCTP e 1U de Klenow DNA polimerase, em um volume final de 25µL. A reação foi incubada a temperatura ambiente por 24 horas e finalizada por aquecimento da mistura a 65°C por 10 minutos. Após a incubação, ao volume final da mistura de reação foi acrescida de 80µL de água mili-Q estéril.

4.11.1.2 Transferência de DNA do gel para membrana de 'nylon'

As amostras de DNA cromossômico dos mutantes foram tratadas com enzimas de restrição (*Bgl*II e *Sal*I), e em seguida submetidos a separação por eletroforese em gel de agarose (0,7%). Depois da corrida eletroforética, o gel foi corado com brometo de etídio, visualizado em transiluminador de luz ultravioleta e registrado por fotografia. Em seguida, o gel foi tratado com uma solução de depurinação (HCl 0,25 mol/L) por 5 min a temperatura ambiente, lavado com água destilada por 2 minutos e tratado com solução desnaturante (NaOH 0,5 mol/L, NaCl 1,5 mol/L) por 30 minutos. O gel foi novamente lavado com água destilada por 2 min e depois tratado duas vezes com solução neutralizante (Tris-HCl 1 mol/L pH 7,5, NaCl 1,5 mol/L) por 30 min. O gel foi então incubado por 15 min em SSC 20X (NaCl 3 mol/L, citrato de sódio 0,3 mol/L pH 7,0). Após este tratamento, o gel foi colocado sob uma membrana de náilon do mesmo formato e

tamanho do gel, evitando-se a formação de bolhas de ar entre o gel e a membrana. Sobre esta membrana foram colocadas 3 folhas de papel Whatmann 3MM cortadas do mesmo tamanho e, sobre estas, foram colocadas diversas camadas de papel absorvente. O sistema foi mantido por 12-16 h para transferência dos fragmentos de DNA para a membrana de náilon por capilaridade. Após este período, a membrana foi cuidadosamente removida e lavada com solução de SSC 1X para remoção de resíduos de agarose. A membrana foi seca a temperatura ambiente e os fragmentos de DNA foram fixados à membrana por irradiação com luz ultravioleta (302 nm) por 5 min.

4.11.1.3 Hibridização

A hibridização do DNA foi realizada conforme Sambrook et al. (1989).

As membranas de náilon contendo os fragmentos de DNA preparadas conforme item anterior, foram colocadas em tubos de hibridização contendo 10 mL de solução pré-hibridização (tampão fosfato pH 8,0, 2% de SDS e 20 µg/mL de DNA de esperma de salmão sonificado e desnaturado) e incubados por 4h a 65°C em forno de hibridização Hybaid sob agitação rotatória. O DNA sonda marcado foi desnaturado em banho fervente por 5 min e resfriado rapidamente em banho de gelo, em seguida, foi adicionado ao tubo contendo a solução de pré-hibridização. As membranas foram novamente incubadas por 24 h a 65°C. Após a incubação, foram lavadas uma vez com solução SSC 2X contendo SDS 0,1% por 15 min a 65°C. As membranas foram colocadas em suporte plástico e adicionado tampão de lavagem para mantê-las úmidas. Em seguida foram cobertas e seladas com filme plástico e autoradiografadas com filme Kodak X-Omat a -70°C utilizando uma placa amplificadora (Dupont, Cronx Hi-Plus), durante 24 h quando o filme foi revelado e analisado.

4.12. ATIVIDADE DA NITROGENASE

A atividade da nitrogenase foi determinada pela redução de acetileno (C_2H_2) a etileno (C_2H_4), pela enzima nitrogenase. O C_2H_4 pode ser mensurado com precisão por cromatografia gasosa. A atividade foi determinada em células crescidas em meio NFbHP semi-sólido contendo glutamato (0,5 mmol/L) (Pedrosa e Yates, 1984).

Cerca de 30 a 50 µL de pré-inóculo das estirpes, cultivadas em meio líquido com A_{600nm} de 1,0, foram inoculadas em frasco de 10 mL contendo 4 mL de meio NFbHP

malato semi-sólido, suplementado com 0,5 mmol/L de glutamato como fonte de nitrogênio. Após 24-30 h de incubação a 30°C ou após a formação de película de crescimento próxima à superfície do meio, os frascos foram vedados com rolha de borracha e injetados com acetileno até concentração final de 10% da fase gasosa. Após um período de uma hora de incubação a 30°C, alíquotas de 0,5 mL foram retiradas para determinação de etileno formado.

A determinação de etileno formado foi realizada por cromatografia gasosa em um cromatógrafo Varian modelo star 3400cx. A temperatura de injeção foi 140°C, da coluna foi 130°C e a do detector foi 200°C. Para análise quantitativa de etileno formado foi utilizada uma mistura de etileno padrão que em 0,5 mL continha a 2,16 nmol de etileno, a 1 atm e 25°C.

A atividade específica da nitrogenase foi expressa em nanomoles de etileno formado por minuto e por miligrama de proteína.

4.13 DOSAGEM DE PROTEÍNAS

A dosagem de proteínas foi realizada pelo método de Bradford (1970). O sistema de reação continha 1 mL da solução de corante (solução de Bradford – 100 mg/L Coomassie Brilliant Blue G-250, 95% etanol e 85% ácido fosfórico) e 100µL da amostra de célula, que foram previamente lisadas pela adição de uma solução de NaOH (0,5 mol/L) e incubação mínima de 30 min à temperatura ambiente. Essa mistura foi homogeneizada e a absorbância a 595nm determinada. A curva padrão foi feita utilizando albumina de soro bovina como padrão, preparada no mesmo tampão das proteínas a serem quantificadas. A falha de linearidade deste método foi de 1 a 10µg de proteína.

4.14 ANÁLISE DAS SEQUÊNCIAS

As comparações de sequência de proteínas foram feitas utilizando o programa ClustalW (<http://www.ebi.ac.uk/clustalw/>) e busca no banco de dados GeneBank foram feitas utilizando o programa Blast (<http://www.ncbi.nih.gov/BLAST/>). As prováveis ORFS foram identificadas pelo Programa Frameplot (<http://www0.nih.gov/~jun/cgi-bin/frameplot.pl>).

5. RESULTADOS

5.1 ANÁLISE DAS SEQUÊNCIAS

A seqüências dos genes *fixNOQP* foram obtidas a partir do banco de dados do Programa GENOPAR (www.genopar.org) e possui 3176 nucleotídeos, que aparentemente fazem parte de um operon contendo ainda os genes *fixI*, *fixS*, *fixG* e *fixH*. Assim uma região de 12099 pb foi selecionada e analisada utilizando o programa Blast (<http://www.ncbi.nih.gov/BLAST/>). Regiões que codificam para uma proteína são ricas em G e C. No gráfico essas regiões aparecem em picos ou acima do gráfico forma de uma linha reta. Possíveis códons de iniciação são indicados por uma flecha nas linhas retas (ATG ou GTG). O códon de parada (TGA) é mostrado por um traço vertical na linha reta. São mostradas as três fases de leitura e estas são diferenciadas pelas cores vermelha (primeira fase), verde (segunda fase) e azul (terceira fase).

O alinhamento das seqüências de aminoácidos das prováveis orfs foi feito utilizando o programa Clustal W (<http://www.ebi.ac.uk/clustalw/>).

As prováveis regiões codificadoras de proteína foram identificadas pelo programa FramePlot (<http://www0.nih.go.jp/~jun/cgi-bin/frameplot.pl>). No total foram identificadas dez ORFs (operon reading fase) nas 12099 pb analisadas.

FIGURA 3a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA HIPOTÉTICA LOCALIZADA A MONTANTE DO GENE *fixI*

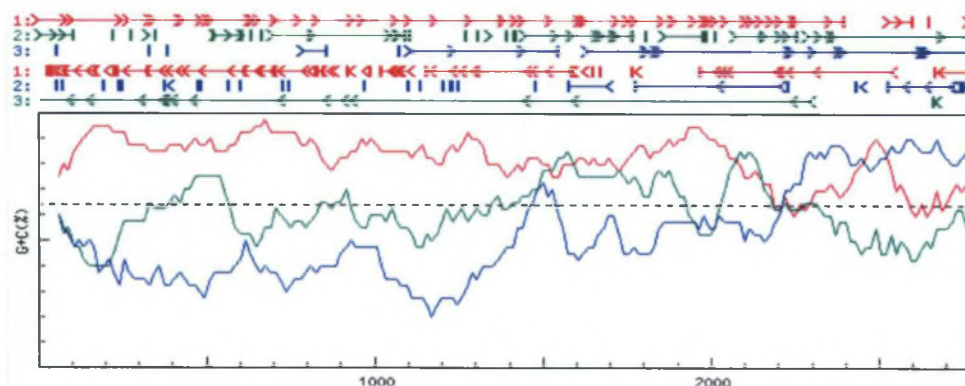


FIGURA 3a - As cores do gráfico representam as fases de leitura, a primeira fase é representada pela cor vermelha, a cor verde indica a segunda fase de leitura e a terceira fase é representada pela cor azul. Acima dos gráficos as possíveis orfs são indicadas pelas linhas, onde as flechas são possíveis códons de iniciação e as barras verticais indicam possíveis códons de parada. Esta proteína hipotética tem sua provável região codificadora iniciando no aminoácido 1390 e o códon de parada na posição 2244, situando-se na primeira fase de leitura.

FIGURA 3b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA HIPOTÉTICA LOCALIZADA A MONTANTE DO GENE *fixI*

| | | |
|------------------|--|-----|
| Hip_Hseropedicae | ---VNLLPIFLVGLMGSVHCIGMCGGIVGALSSAAPVRPTPTPAAPLLSVAAPGRNDGAR | 57 |
| Hip_Daromatica | MPDSGYLALFLVGLLGTHCVGMCGGIVGALSMGGQAR----- | 38 |
| | * | |
| Hip_Hseropedicae | LLRIIPIHAGGVQVQAMAQDLVRVLCYNLGRLLSSYALAGALAGGIAAGLLRGADVLGWL | 117 |
| Hip_Daromatica | ---WSMH-----LAYNGGRILSYAAAGATAGALGAASMG---LEGQ | 73 |
| | * | |
| Hip_Hseropedicae | APAQRVAYLITNIVLVLLGLYLTQWWPALARLEQWGSALWVRVRPLAARLVVDTPAKAL | 177 |
| Hip_Daromatica | VPARLILYFVANMLVAMGLYLLGVTKALAFTERAGQSLWRRQLPLTRRFLPVRVSAQAF | 133 |
| | * | |
| Hip_Hseropedicae | LLGSLWGWLP CGMVYSALLTALMAGSAMQGALTMLAFGAGTLPVLLAAGLSGARLRQLAQ | 237 |
| Hip_Daromatica | PLGVWGWLP CGLVYSALATALSTGSAGRGALTMLAFGLGTLPNLLLAGIVLARLNGFVR | 193 |
| | * | |
| Hip_Hseropedicae | RPAVRLAAGVVVLAFGIMLLRGAIEIGMGIAARGWIDVFCISPAHGG | 284 |
| Hip_Daromatica | RPIVRTLSGLLVLFGLYGLL-----GLMRLSGGL----- | 223 |
| | * | |

FIGURA 3b - O símbolo "***" indica que os aminoácidos são idênticos em todas as seqüências, o símbolo "*" mostra que os aminoácidos possuem características semelhantes porem ainda assim são diferentes. O símbolo "." indica aminoácidos semi-conservados. As letras

representadas pela cor vermelha são aminoácidos hidrofóbicos, os azuis são ácidos, os de cor magenta são os básicos. Os aminoácidos representados pela cor verde tem características hidroxil – amino – básico – Q, outros tipos estão representados pela cor cinza.

Como mostram as figuras 3a e 3b a primeira orf inicia na posição 1390 e possui no total 284 aminoácidos. Ela foi classificada como uma proteína hipotética, porém é bem conservada em bactérias, geralmente ocorrendo a jusante de genes codificadores da proteína citocromo *cbb₃* oxidase. Na subclasse Epsilon Proteobacteria uma sequência homóloga foi encontrada a montante. Esta é provavelmente uma proteína transmembrana. A mesma proteína de *Dechloromonas aromatica* apresentou 44% de identidade e 56% de similaridade com *H. seropedicae*.

FIGURA 4a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA FixI

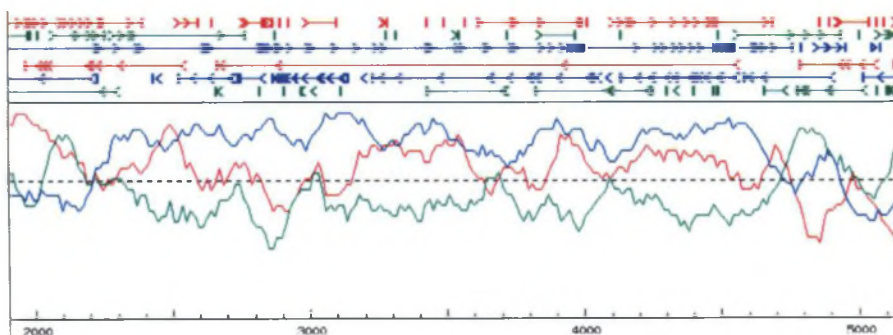


FIGURA 4a – As cores do gráfico são descritas conforme legenda da figura 3b. A proteína FixI tem sua provável região codificadora iniciando no aminoácido 2313 e o códon de parada na posição 4763, situando-se na terceira fase de leitura.

FIGURA 4b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FixI

```

fixI_Hseropedicae      -----VPAG-----QAWTLEIGGARRELCCV 21
fixI_Rsolanacearum     MTTTLAPILAGTPSAGSISSGRTDDARASHVTCYHCGAPLDRAAALHAEMAGASRDFCCA 60
                        ::*      :.:.**::**

fixI_Hseropedicae      GCQSVARLIVDSGCEDFYLRRTAPSARVDPEQLPPELALLEQPQAPDQAH---ADAS 78
fixI_Rsolanacearum     GCQTIAMTLHASGMGHVYDGPIDAFARPIDGEARTEAEAAWATYDLPVMRERFVRTREDGA 120
                        ***::* : ** : : : : * : : : : * : : : : * : : : :

fixI_Hseropedicae      AELVLSIDGLRCSACVWLIEKYLARLPQMMAEMNVATSRLHLRRDPALCSTATILRGLR 138
fixI_Rsolanacearum     EEVSLTLVGMRCACVWLIERALARVPGVRESIVNYATERVRIVSEAGAVRLADLFAAIA 180
                        * : * : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      GLGYTAYFPDPLRQGEQARRASRLFRQLFIAGLSMMQVMYAVPVYMTHEGIDPDMMSL 198
fixI_Rsolanacearum     DVGYEAWPDQPSLRRTAEARERRHLIRLGIAMLGMMQVMYAWPVYLHGGSI PADQVAL 240
                        : : * : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      MRWASLFLTIPAVFYSAFPFFTGAAGLRARAPGMDLPVAIGIAAFLASAIATWRGEGE 258
fixI_Rsolanacearum     MQWASLALTTPVVLFSAGPIFRSAWQQVRHAHVGMDDVPVALGIGAAFAASVAATVRGHGA 300
                        * : * : * : * : * : * : * : * : * : * : * : * : * : * :

fixI_Hseropedicae      IWFDSVSMFIFLLLSGRYLEAAARRKSASALERMQQAFPASALLAPGYPEQRETTLVVAA 318
fixI_Rsolanacearum     TYFDSVTMFVAFLLAARYLELRARQSATSGTEALVRQLPATCRRIVDGAESTQT--VPVA 358
                        : : : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      QLREGDVILARPGDTIAADACLLGAGEFDLALLSGESRPQSFQIGQEAPGGAVNLSQPV 378
fixI_Rsolanacearum     ALQAGDCVEVRAGEILPADGVIERGTTEVDESLLSGESLPQPRGVGAAGVLAGSYNVAIAI 418
                        : : * : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      FLRVVRATGESTLAALTRLAEQAGQGKPALAQWADVVASRFVALLLLAGATFLAWQLID 438
fixI_Rsolanacearum     RLRVHRVGAQTRLAAIVDLLDLALTDKPRMAELADRVAGRFBVVLGLWAALTACAWWID 478
                        * : * : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      PTRAWSTAI AVLVS CPCALSLATPSALAAATDRLLRAGTLIVRGNVLETQRADTIVFD 498
fixI_Rsolanacearum     PGRMFAVTVAVLVVSCPCALSLATPSALAAAGALARRGVLVTRGHATIESLAATDVLLD 538
                        * : : : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      KTGTLTQGRPQLT--ALWSDVDCQRSLAIAAAMERGS LHP LAKALVEEGARRQVDALEVE 556
fixI_Rsolanacearum     KTGTLTQGRRLRLS IETFDLDAQRCLALACALEQAENHP I AQSLRAA--VDAPAPSM AVD 597
                        * : * : * : * : * : * : * : * : * : * : * : * : * : * :

fixI_Hseropedicae      QLTSVTGAGMQGMIEGDLYRIGSRSFVAGLAGTALPPVLLQSAASGAG-----SVYL 608
fixI_Rsolanacearum     EVLVNPGQGVRRARWQAQDLRLGTQAFAGFTGQLPRTSVHVGSEPAEGMPPTAACTAIWL 657
                        : : * : * : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      GSSDGWLARFDLADALRTDAVATVAAFRALGLRTILLSGDQPEVCAEVAAAAGISEVQAA 668
fixI_Rsolanacearum     GRDGRPLARFVLADTPRADAACLDALRGQGLRLHLVSGDAPETVRWWAARLGIHAHEGG 717
                        * : : : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      CTPQGKLDYVRNLQRDGA VVAIVGDGINDAAML RAGDV SFAMGKAALAQVSA DAVIMSD 728
fixI_Rsolanacearum     ASPEDKRAYVRALQAGGARVLAVGDGINDA PLLAQAVS I AIGTGAPLAQAGADAILTEP 777
                        : * : : * : * : * : * : * : * : * : * : * : * : * : * :

fixI_Hseropedicae      RLQAVADCAAMAQRTMHIVRQNLVWAATYNFLAIPAAAFGLLDPWMSAVGMSLSLLLVVG 788
fixI_Rsolanacearum     RLPAIGEAVSIGRRTL RVVRQNLGWAFAYNAVSIPLATLGWLSPLEAGIGMSLSLLLVAL 837
                        * : : : : : : : : : : : : : : : : : : : : : : : : : :

fixI_Hseropedicae      NALRLSRRQVRSAAEAQPAAVRPALAGAA 817
fixI_Rsolanacearum     NAWRLSR-----AA----- 846
                        ** **** **

```

FIGURA 4b – Legenda conforme figura 3b.

A sequência de aminoácidos da segunda orf analisada possui 44% de identidade com a proteína FixI de *Ralstonia solanacearum* e 43% com a de *Chromobacterium violaceum* e portanto o gene foi denominado *fixI* de *H. seropedicae*. Tem início no nucleotídeo 2313 e possui um total de 817 aminoácidos. Esta proteína é uma provável ATPase do tipo P (figura 4a e 4b).

FIGURA 5a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA FixS

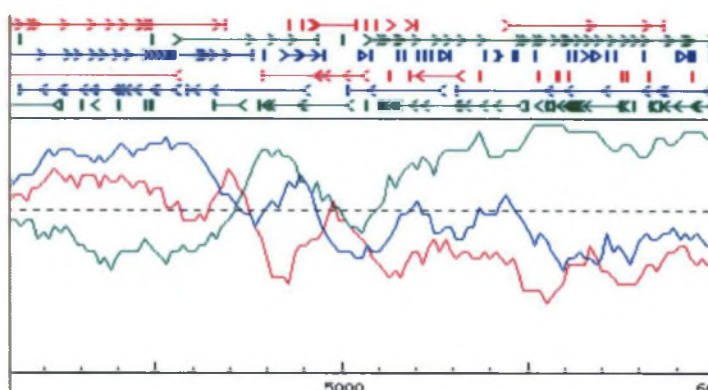


FIGURA 5a – As cores do gráfico são descritas conforme legenda da figura 3b. A proteína FixS tem sua provável região codificadora iniciando no aminoácido 4763 e o códon de parada na posição 4939, situando-se na segunda fase de leitura.

FIGURA 5b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FixS

```

fixS_Hseropedicae   AGGSGLM EALYLLIPLSTLLVFLAIWVFFRASDSGQFDDLEG PAMRILHDDDAAVPQEPS 60
fixS_Cviolaceum     -----MESLYLLIPLSLVLVFI GAIFFWATRSGQFDDMQGPAHRILMDDDATHEDDKS 54
fixS_Mloti          -----MTILVYLM PVALFLGALGTGFLWALKSGQYEDLDGAAERILRDDKPER----- 49
                    *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
fixS_Hseropedicae   GGA 63
fixS_Cviolaceum     SG- 56
fixS_Mloti          ---

```

FIGURA 5b – Legenda conforme figura 3b.


```

fixN_Hseropedicae  TDPIKFLVVSLSFYGMSTFEGPMAIKTVNALSHYTDWTIGHVHSGALGWVGFVTMGAM 360
fixN_Cviolaceum    TDPVLKFLVVSLSFYGMSTFEGPMMSIKTVNALSHYTDWTIGHVHSGALGWVGFITIGSM 359
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

fixN_Hseropedicae  YYLIPRLSGKTQMWSKDLIEIHFVWATIGIVLYIAAMWIAGVMQGLMWRAVNADGTLTYT 420
fixN_Cviolaceum    YYLIPRLFGRETMSVSKLIEAHFWLATVGVVLYIAALWISGVMQGLMWRAVLPDGTLYA 419
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

fixN_Hseropedicae  FVESVKATYPYYVIRLMGGLMYLSGMLVMAYNTWRTMRGTELAVAPIPAVTNAAH 475
fixN_Cviolaceum    FVEAVKATYPYHTVRLGGLYLSGMLLMAYNTYRTVVAGRAVDKIPAITAEAH 474
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :
                    * * * * * : * * * * * : * * * * * : * * * * * : * * * * * :

```

FIGURA 6b – Legenda conforme figura 3b.

A próxima orf inicia 145 pb a jusante de *fixS* e codifica para a primeira subunidade da proteína citocromo *cbb₃* oxidase, a proteína FixN. FixN possui 475 aminoácidos e sequência bastante conservada entre diversos organismos (figuras 6a e 6b). A proteína FixN de *H. seropedicae* apresenta 79% de identidade e 88% de similaridade com a proteína ortóloga de *C. violaceum*. Apresenta também 64% de identidade e 81% de similaridade com a de *M. loti*.

FIGURA 7a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA FixO

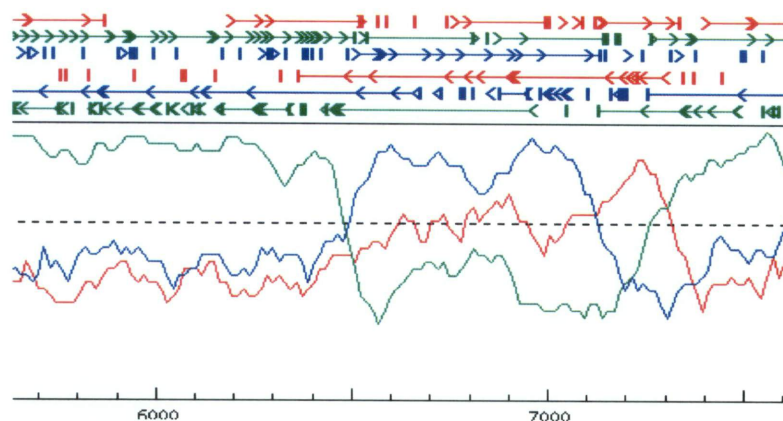


FIGURA 7a – As cores do gráfico são descritas conforme legenda da figura 3b. A proteína FixO tem sua provável região codificadora iniciando no aminoácido 6525 e o códon de parada na posição 7139, situando-se na terceira fase de leitura.

FIGURA 7b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FixO

```

fixO_Hseropedicae  -----MKFSHEWIEKNPWLLIGLVLLVVSIGGLAEIVPLFFQKSTTEPIAGLKPY 51
fixO_Terythraeum   MSDHNNSAPKSFSEKVEETNNFLLIVLTLLVVAVGGMVEIVPLFFQKSTTEAVAGLKPYT 60
                  : * * * : * : * * : * : * : * * * * * * * : * * * * *
fixO_Hseropedicae  ALRLAGRDVYIREGCYNCHSQMIRPLRAETERYGHYSVAGESVYDHPFQWGSKRTGPDLA 111
fixO_Terythraeum   PLQLMGRDVYLRGCYNCHSQMIRPFRAETMRVGHYSVAGEFVYDHPFQWGSKRTGPDH 120
                  : * : * * * * : * * * * * * * * * * * * * * * * * * * *
fixO_Hseropedicae  RVGTRYSDewHRAHLHNPRDVPESNMPAYPWLEQAKLDNYDIVSRMKALKRIGDPYTEE 171
fixO_Terythraeum   RVGGKYSDEWHRIHLNNPRDVPESNMPAYTWLEKNKVDDTVVAQHMSALRKVGVPYSDE 180
                  * * * : * * * * * * : * * * * * * * * * * * * : * : : * : * * * *
fixO_Hseropedicae  DIKNAPAEVLVGKTEQDALIAYLQGLGVLIKAER 204
fixO_Terythraeum   EISGAEEQVKGKTEMDAVIAYLQVLGRSVK--- 210
                  : * . . : : * * * * * : * * * * * * * : *

```

FIGURA 7b – Legenda conforme figura 3b.

Dezessete pares de bases a jusante começa o gene *fixO* que codifica a subunidade II da citocromo *cbb₃* oxidase esta proteína possui 204 aminoácidos e uma identidade de 70% com a de *Trichodesmium erythraeum*. Sua sequência parece ser bem conservada em relação a de outros organismos como mostram as figuras 7a e 7b.

FIGURA 8a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA FixQ

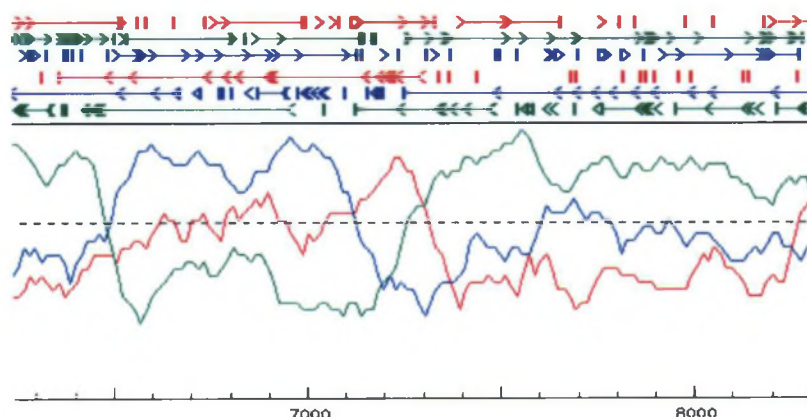


FIGURA 8a – As cores do gráfico são descritas conforme legenda da figura 3b. A proteína FixQ tem sua provável região codificadora iniciando no aminoácido 7153 e o códon de parada na posição 7338, situando-se na primeira fase de leitura.

FIGURA 9b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FixP

```

fixP                                MADFTNGFWNIWIIIVLTVLG-----AGSTTGHVWDEDLTELNNPLP 41
fixP_Polaromonassp.                MSDFTNFWSLSLYVAALTLLGLVACLVLWVLTARKRIVSSADNTTGHVWDGDLREMNPNMP 60
                                     *:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
fixP                                RWWWMLWFYLTIFFSIGYLVAYPGLGNLPGLTGWSTGEHIVGCTLLWLQQSNWVKVKKEE 101
fixP_Polaromonassp.                RWWWLWFLVLTIVFSLAYLVAYPGLGAYQQQLGWSTRGEYAV----- 101
                                     **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
fixP                                KADADVKAABEAKYGPLFNHYLKQDLKVVAGDSQAHAIGERLFLTYCAQCHGSDARGNKG 161
fixP_Polaromonassp.                ----EVERANKELGPLYSQFTAKPTEELAGDANAMAVGERLFMNNCAQCHGSDARGSKG 157
                                     :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:
fixP                                PNLTDNDWLHGGTPEIIKETIMKGRHGVMPMAAAVGTDKDVDNVANYVLSLSDSTHDP 221
fixP_Polaromonassp.                PNLTDADWLHGGTPDKITETLTRGRQQQMPPMAAAVGTDPDVKNLANVYVLSLGS SPHDSV 217
                                     **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
fixP                                KAELGKSKFAACMACHGPGGKGNQALGAPNLSDKIWLYGGSIDTIKETINKGRDNTMPAF 281
fixP_Polaromonassp.                RGALGKSKFGACAACHGVVDGKGNPALGAPNLADDVWLHWGEQAIINMINKGKTSEMPAQ 277
                                     :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
fixP                                GEFLGEPKVHVLAAYVWSLSNKPNSAAAK 310
fixP_Polaromonassp.                LKGLTEAQLHVLTAIYWGMSNKPLPAKP- 305
                                     *:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

FIGURA 9b – Legenda conforme figura 3b.

A próxima orf codifica para a terceira subunidade desta proteína, a proteína FixP de *H. seropedicae*. O gene *fixP* inicia no nucleotídeo 7280 e apresenta 310 códon. A proteína FixP de *Polaromonas sp.* apresenta 57% de identidade com a de *H. seropedicae* (figuras 9a e 9b).

FIGURA 10a – IDENTIFICAÇÃO DAS POSSÍVEIS REGIÕES CODIFICADORAS DAS PROTEÍNAS FixG E FixH

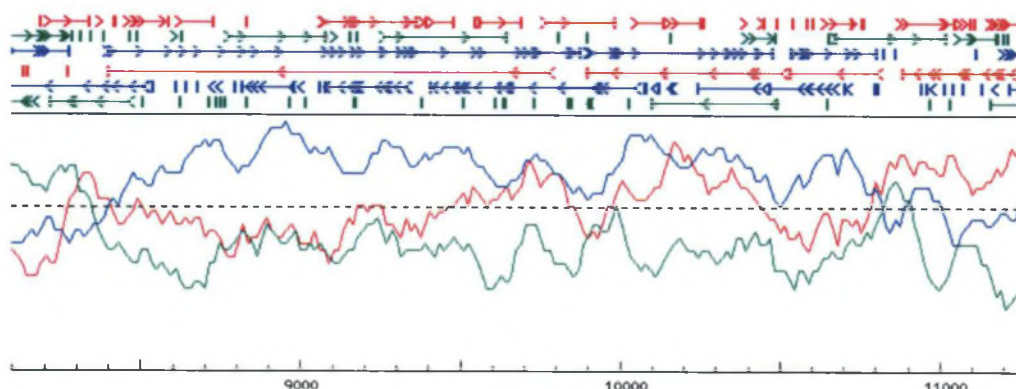


FIGURA 10a – As cores do gráfico são descritas conforme legenda da figura 3b. As proteínas FixG e FixH têm suas prováveis regiões codificadoras iniciando no aminoácido 8406 e 9909 respectivamente e o códon de parada na posição 9875 e 10478, ambas na terceira fase de leitura.

FIGURA 10b - ALINHAMENTOS DAS SEQUÊNCIAS DE AMINOÁCIDOS DEDUZIDAS DAS PROTEÍNAS FixG E FixH

| | | |
|-------------------------|--|-----|
| fixG_Hseropedicae | MEKAVMKPRQDKEGDPPDTGDSKTGKDKEYAIVIRMYAAREQIYPREIQGR | 50 |
| fixG_Pnaphthalenivorans | -----MTTPKKVIPITIAQAADAEMISLYAAHEKIYPRSVSGF | 37 |
| | * * : : * | : |
| fixG_Hseropedicae | FASLRWLCVF LTQL VF YGLPWINWNERNR QAVLFDLASRK FY LFG LVLPQD | 100 |
| fixG_Pnaphthalenivorans | FASWRWG VVL T QL VF Y GLPWIE W GQRQ AVLF DL ASRRFYIFGLV LPQD | 87 |
| | : : | * |
| fixG_Hseropedicae | FIWL AALL IICAFSLFLTAVAGRVWCYGSCPQT VYTE IFMWIERRIE GN | 150 |
| fixG_Pnaphthalenivorans | FIYL TGILVISAL SL FL TAVAGR LWCGYAC P QT VTYTEIFLWI ERKM EGD | 137 |
| | : : : : * : * | : |
| fixG_Hseropedicae | RSARMRLDRQPWS F DKL WRKS AKHLAWGAVALWTGISFVG YFSPI RDL LP | 200 |
| fixG_Pnaphthalenivorans | RAVQMRRDGQPWS PG RL WRKG GH LL WLAVA VAL WTGF TF VG YFT PI K VL GM | 187 |
| | * : : * * * | : |
| fixG_Hseropedicae | EIS GF ALGPWE SF WI LF Y GFATYN AGWMREQVC KYMCPYARFQS AMF DR | 250 |
| fixG_Pnaphthalenivorans | EFVQMS MSS WE AF WF FF Y GFATYN AGFM REQVK HKMC PY ARF QSAMF DK | 237 |
| | * : :: : * | : |
| fixG_Hseropedicae | D S LIITYDVARGE PRMP AA KA KL DTGAR-AGDCIDCTMCVQCPTGTIDI | 299 |
| fixG_Pnaphthalenivorans | D TL IV TYDAER GD PRG--ARS RKADPAAL NLGS CV DCN LC V QC PT GT IDI | 285 |
| | * : * : * | : |
| fixG_Hseropedicae | RQG L QYM CIGCA ACVDAC DSVM DIKR PRGL IRYSTENAVEGGFSTA EI R | 349 |
| fixG_Pnaphthalenivorans | RKGLQEYC IG CGACA DV CN TV MD KMGYER GL VKYT TENAM KN KW TAQT W | 335 |
| | * : * : * | : |


```

fixG_Hseropedicae      RRLVVRPRILIIYGAILGAVIALFAGSLWVRTPLKLDVIRDRGSMGREVEEG 399
fixG_Pnaphthalenivorans RHVLRPRILIIYTAILLGIVMAMLVSLTLRTPFKVNVVRDRGVMARIVSGG 385
                        *:::***** ** : : * * : * * : * : * : * : * : * : *
fixG_Hseropedicae      I IENVYRLQIINTDERGHRVLRVAKGLAGLSVDPATPIEVAATQTVSVVPV 449
fixG_Pnaphthalenivorans NIENVYQLQVMNATEADQRYKIRVTGLPGLAITSSEVVMMVESTQARRLAV 435
                        *:::***** : : * * : * * : * : * : * : * : * : *
fixG_Hseropedicae      RVRAPHGAGEVGSNKIRIELEAEDQPALNVSEKAVFLVPRR 490
fixG_Pnaphthalenivorans RVQAPYEAAAPGTHPIAFEDALDS-AGHLVEKSVFMVPR- 474
                        *::: * : * : * : * : * : * : * : * : * : *

fixH_Hseropedicae      MQTMSPKFASTAAAVPWYRHRWPWLLMAGPAVVVVAGVFTAWLAISRADALVADDYYKQG 60
fixH_Daromática        ---MSNSMTLRSNQPWYKERWPWILMAGPAVVIVAGVITLWLAVVSNGLVTDYYKQG 57
                        * * : : : * * : * * : * * : * * : * * : * * : * *
fixH_Hseropedicae      KAINKDLRRDHEAQLGAAIDIAYPDPAAGVLRGRLQLRQPAPQAQSQLADQASRTLIINL 120
fixH_Daromática        LAVNQQLHREHQAGNLGLHADVMRAGAN--IR-----LLTAADGDAKLP----SALFLKL 106
                        *:::***** * * : * * : * * : * * : * * : * : * : *
fixH_Hseropedicae      VHPTQPSKDRTLQVQPAADGSFSVPLVEMEQRWRLVAEDGSHVWRLHGTWAWPQQRSM 180
fixH_Daromática        AHPTRAGQDQSVQMSSEGQGFYSGKLAAEVTGRWLVSIEDPAGQWRLQGEWQADSEELR 166
                        * * : : * * : * : * * : * * : * * : * * : * * : * : *
fixH_Hseropedicae      MNAEAYAPAE 190
fixH_Daromática        LVAKADK-- 173
                        : * *

```

FIGURA 10b – Legenda conforme figura 3b.

Na maioria dos organismos o gene *fixP* é seguido pelos genes *fixGHIS* (Kahn, 1989). De fato, em *H. seropedicae* os próximos genes são o *fixG* e o *fixH*, porém o *fixI* e o *fixS*, estão localizados a montante de *fixNOQP*. Como mostram as figuras 10a e 10b o gene *fixG* apresenta neste caso 490 códons, e seu produto de tradução 59% de identidade com a proteína *fixG* de *Polaromonas naphthalenivorans*. O produto do gene *fixH*, que é provavelmente uma proteína de membrana com 190 aminoácidos, apresenta 49% de identidade com a de *D. aromática*.

FIGURA 11a – IDENTIFICAÇÃO DA POSSÍVEL REGIÃO CODIFICADORA DA PROTEÍNA FNR

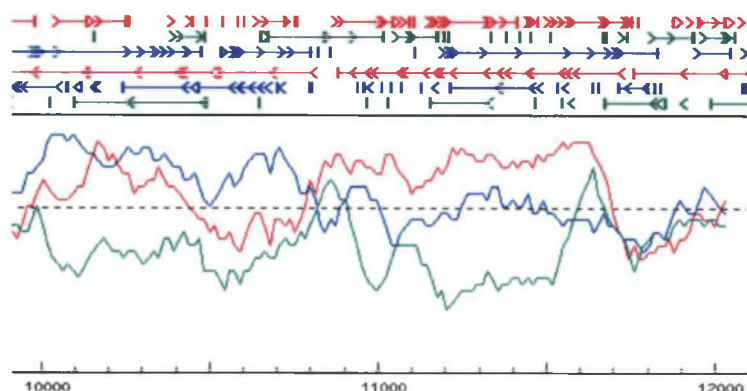


FIGURA 11a – As cores do gráfico são descritas conforme legenda da figura 3b. A proteína FNR tem sua provável região codificadora iniciando no aminoácido 11576 e o códon de parada na posição 10878, situando-se na primeira fase de leitura.

FIGURA 11b - ALINHAMENTO DA SEQUÊNCIA DE AMINOÁCIDOS DEDUZIDA DA PROTEÍNA FNR

```

Fnr_Rgelatinosus  MKRMEREPIRIEALKVACSSCNLRELCLPVGLSDDDLHLDQLVGSRRPVKRGDTLFHAG 60
Fnr_Cviolaceum    MS--EQNHTSLHTLTKISCSSCSLRELCLPVGLNREEMGQLDAVIRQSRRLKRGEYLFRRSG 58
Fnr_Hseropedicae  -----MHQLCLPMGLDESDMKRLDKIIG-RRKVARDDDFLYRIG 37
                  :::::***:***:  :: :***:  :  :  :***:  :  :  :

Fnr_Rgelatinosus  DRFESLYAVRTGFFKTCVSADDGRDQVTGFQMAGELLGLDGISTDRHACDAVALEDSQVC 120
Fnr_Cviolaceum    ESFKSLYAIRTGFFKTCVASQDGREQVTGFLMSGELMGLDGISSSNHSCDAVALEDSEVC 118
Fnr_Hseropedicae  DRFTALYAVRVGHFKTYQENLDGDRQITGFQMPGELLGMDAISTEQHQCDVAVALQDSEVC 97
                  :  :  :***:***:***:  ***  :***:***:***:***:***:***:***:***:***:

Fnr_Rgelatinosus  QIPYSQLEKLSRESTPLQHGFHKIMSREIVRDHGVMLLLGSMRAEERLA AFLNLTLQRLQ 180
Fnr_Cviolaceum    ELPFTRMESLCRDIPSLQHGFYRLMSREIVRDQNVMLLLGNMKAEEERIA AFLNLNLSQRLS 178
Fnr_Hseropedicae  EIPFARLEQLFGQIPHLLRHFHRIMSHETSEQNVIMLLGNMRAEQRF AAFVLNLSRYA 157
                  :::::***:***:  :  :  :***:***:***:  :  :  :***:***:***:***:***:***:***:

Fnr_Rgelatinosus  ARGFSASSLVLRMTREEIGSYLGLKLETVSRTFSKFQDEGLLEVQQRDIRVLDESGLRRL 240
Fnr_Cviolaceum    TRGFAANDFILRMSREEIGSFGLKLETVSRTLSEKFQQGWITVDHKKHIQLVQLAELKLN 238
Fnr_Hseropedicae  ARGYSSTRFQLRMTRQDVGNYLGLTIESISRLISKFRKQGLLAVEQRDVEVVDLAALKRL 217
                  :***:***:  :  :***:***:***:***:***:***:***:***:***:***:***:***:***:

Fnr_Rgelatinosus  VNGPAC----- 246
Fnr_Cviolaceum    ISGCMHADGH----- 248
Fnr_Hseropedicae  AAGVDACTATPTRSTS 233

```

FIGURA 11b – Legenda conforme figura 3b.

A última orf encontrada está na orientação oposta codificada pela fita complementar. Esta orf inicia no nucleotídeo 11576 e termina na posição 10878, seu produto de tradução contém 233 aminoácidos e possui alta similaridade com a proteína Fnr de diversos organismos: 53% de identidade com Fnr de *Rubrivivax gelatinosus* e 50% com a *C. violaceum* (figuras 11a e 11b). Este gene foi portanto denominado gene *fnr* de *H. seropedicae*.

É possível que a proteína Fnr controle a expressão dos genes *fixISNOQPGH* como ocorre em *S. meliloti*, mas isto não pode ser mostrado neste trabalho.

5.2 INSERÇÃO DO TRANSPOSON NO GENE *fixN*

O plasmídeo HS17058H11 obtido do banco de clones do projeto Genopar foi re-sequenciado para confirmar a presença do gene *fixN* de *H. seropedicae*. Com o objetivo de mutagenizar o gene *fixN* foi feita uma reação de inserção do transposon EZ:TN <TET-1> *in vitro*, aonde o transposon (contendo o gene de resistência a tetraciclina) poderia inserir-se em qualquer posição no plasmídeo (figura 12).

FIGURA 12 – ESQUEMA DE INSERÇÃO DO TRANSPOSON EZ:TN <TET-1> EM PLASMÍDEO

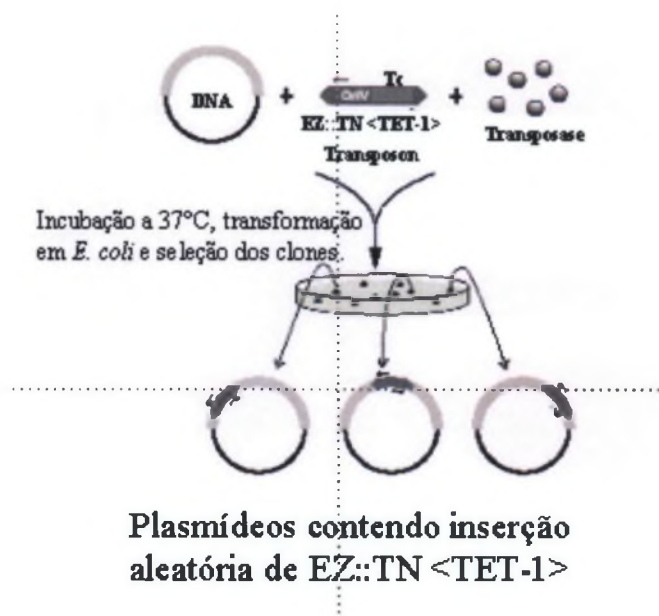


FIGURA 12 - Esquema da inserção do transposon EZ:TN <TET-1> em plasmídeo. Fonte: catálogo da EPICENTRE.

5.3 SELEÇÃO DOS MUTANTES PLASMIDIAIS E CROMOSOMAIS

Após a reação de transposição, os seus produtos foram transformados em células de *E. coli* e cerca de 200 colônias resistentes a marca do transposon (tetraciclina) e do vetor (ampicilina) foram obtidas. Doze colônias tiveram o seu plasmídeo analisado e em quatro deles o transposon tinha entrado no interior do gene *fixN*.

A mutagênese do gene *fixN* do cromossomo de *H. seropedicae* ocorreu por recombinação homóloga após inserção dos plasmídeos mutagenizados. Neste processo o gene mutagenizado presente no plasmídeo recombinau com o gene do cromossomo. Esse tipo de processo pode gerar recombinação simples com mutantes onde o plasmídeo inteiro é inserido no cromossomo bacteriano, ou pode ocorrer recombinação dupla onde ocorre dois eventos de quebra e ligação com a inversão geral do transposon (figura 13).

FIGURA 13 – ESQUEMA DE INSERÇÃO DO TRANSPOSON EZ:TN <TET-1> NO GENE *fixN* CROMOSSOMAL DE *H. seropedicae* SMR1

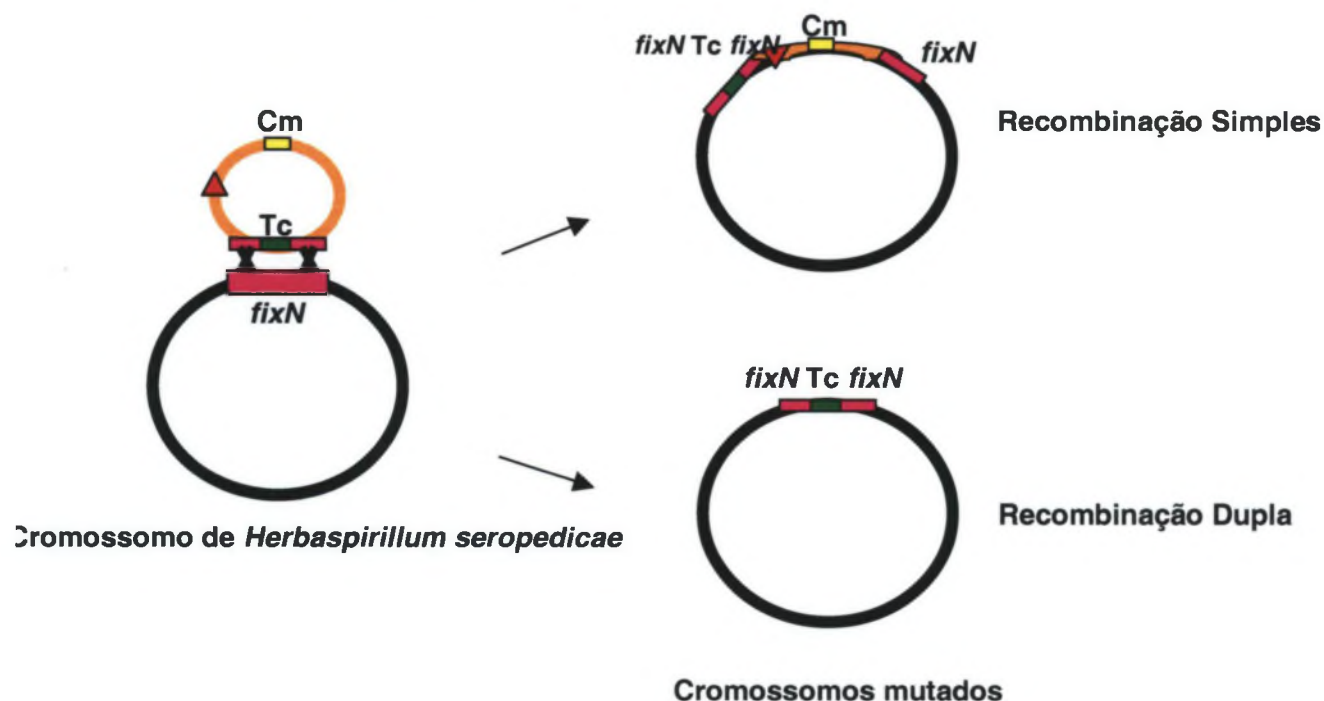


FIGURA 13 - Esquema de construção de mutantes de *H. seropedicae* por recombinação simples e dupla.

Para a diferenciação destes mutantes foi utilizada a seguinte estratégia. Ao plasmídeo contendo o gene *fixN* mutagenizado foi inserido, fora da região homóloga, um cassete que confere resistência ao antibiótico cloranfenicol. Após a transformação no *H. seropedicae* com o plasmídeo mutagênico foram selecionadas colônias resistentes ao antibiótico tetraciclina. Foram obtidas quinhentas colônias resistentes a tetraciclina. Cem colônias foram testadas para resistência ao cloranfenicol e, destas, apenas vinte e duas não foram resistentes indicando a perda do marcador plasmidial. As vinte e duas resistentes apenas em tetraciclina, provavelmente são resultados de dupla recombinação e possuem o transposon integrado ao gene *fixN* cromossomal. Uma estirpe mutante foi isolada e denominada RAMM21.

5.4 ANÁLISE E CONFIRMAÇÃO DOS MUTANTES

A digestão do DNA genômico da estirpe selvagem com a enzima *Bgl*II resulta em um fragmento de 5009 pb, que corta o gene *fixN*. Se houve a inserção do transposon, esse fragmento aumenta para 6683 pb. A figura 14 apresenta o local de inserção e as posições dos sítios da enzima *Bgl*II na região do gene *fixN* de *H. seropedicae*.

FIGURA 14 – LOCALIZAÇÃO DO TRANSPOSON EZ:TN <TET-1> NO GENE *fixN* E REGIÕES DE RESTRIÇÃO DA ENZIMA *Bgl*II

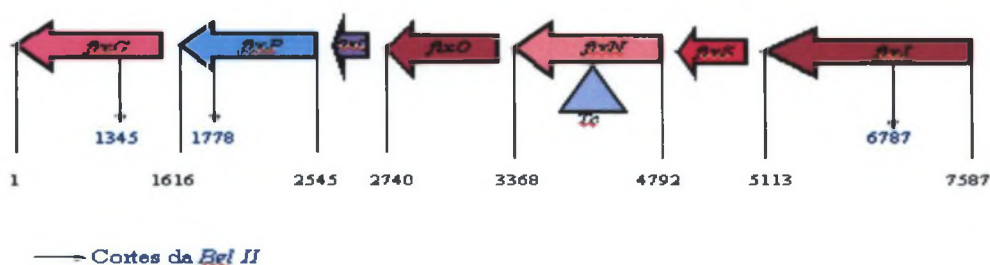


FIGURA 14 - Organização dos genes *fixISNOQPGH* de *H. seropedicae*. Os números em azul indicam os sítios da enzima *Bgl*II, usada para análise de restrição. O triângulo mostra o ponto de inserção do transposon EZ:TN <TET-1>.

Para confirmar a inserção e localização do transposon na estirpe mutante RAMM21, o DNA cromossomal desta estirpe foi isolado, digerido com a enzima *Bgl*II e os fragmentos resultantes separados em eletroforese em gel de agarose. Em seguida o DNA foi transferido para uma membrana de náilon e hibridizado com uma sonda, contendo *fixN* de *H. seropedicae*, o transposon e o plasmídeo vetor. A figura 15 mostra a autoradiografia. Na linha 2 (estirpe selvagem) um único fragmento de aproximadamente 5 Kb contendo o gene *fixN* hibridizou. Na linha 3 (estirpe mutante) o fragmento que hibridiza possui cerca de 6,5 Kb, indicando a inserção do transposon (1,6 Kb) no gene *fixN*. Estes resultados confirmaram que a estirpe RAMM21 possui uma inserção do transposon EZ::TN <TET-1> no gene *fixN* de *H. seropedicae*, resultante de um evento de dupla recombinação.

FIGURA 15 – PERFIL DE HIBRIDIZAÇÃO DAS ESTIRPES SELVAGEM SMR1 E DO MUTANTE *fixN* RAMM21

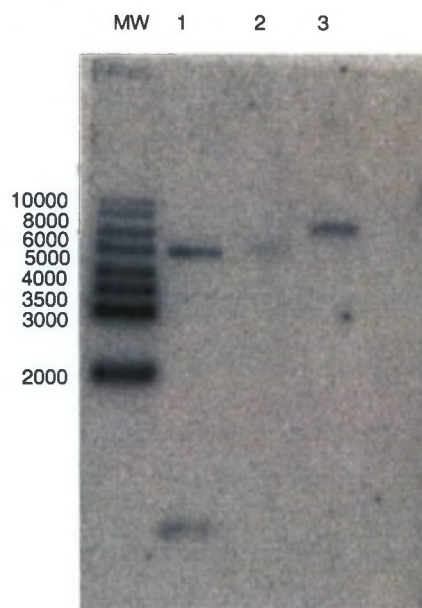


FIGURA 15 - Hibridização do DNA cromossomal da estirpe selvagem Smr1 e o mutante *fixN* RAMM21 de *H. seropedicae*. Autorradiografia da hibridização com sonda contendo o gene *fixN*, transposon EZ:TN <TET-1> e o vetor. MW – marcador molecular. Linha 1 – plasmídeo pRAMM1N digerido com *SalI*. Linhas 2 e 3 mostram o DNA cromossomal digeridos com *BglII* da estirpe selvagem Smr1 e o mutante RAMM21 respectivamente.

5.5 DETERMINAÇÃO DO FENÓTIPO

A estirpe mutante *fixN* de *H. seropedicae* apresenta uma redução de aproximadamente 50% na atividade da nitrogenase em relação a estirpe selvagem Smr1. A Figura 6 mostra esta comparação.

Assim como em *H. seropedicae*, mutantes *fixN* de *Azorhizobium caulinodans* também tiveram a atividade da nitrogenase de 40 a 50% de queda com relação a estirpe selvagem (Mandon et al., 1994). Mutantes *fixN* de *B. japonicum* não são capazes de fixar nitrogênio em condições baixas de oxigênio ou em anaerobiose (Lopez et al., 2001).

FIGURA 16 – ATIVIDADE DA NITROGENASE DAS ESTIRPES SELVAGEM SMR1 E MUTANTE RAMM21.

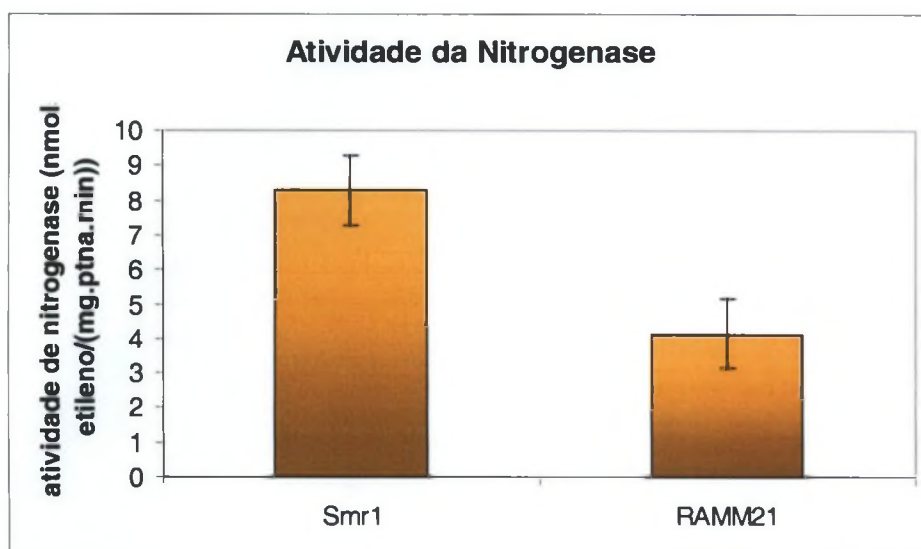


FIGURA 16 - Atividade de nitrogenase de *H. seropedicae* estirpe selvagem Smr1 e a estirpe mutante RAMM21. Atividade da nitrogenase em nmol de etileno por miligrama de proteína por minuto.

6. CONCLUSÃO

Mutantes *fixN* de *H. seropedicae* mostraram que o produto deste gene pode estar envolvido no transporte de elétrons influenciando a atividade da nitrogenase neste organismo.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ANTHAMATTEN, D.; HENNECKE, H. The regulatory status of the *fixL*-like and *fixJ*-like genes in *Bradyrhizobium japonicum* may be different from that in *Rhizobium meliloti*. **Mol. Gen. Genet.**, v. 225, p. 38 - 48, 1991.

ANTHAMATTEN, D.; SCERB, B.; HENNECKE, H. Characterization of a *fixLJ*-regulated *Bradyrhizobium japonicum* gene sharing similarity with the *Escherichia coli fnr* and *Rhizobium meliloti fixK* genes. **J. Bacteriol.**, v. 174 (7), p. 2111 - 2120, 1992.

BALDANI, J. I.; BALDANI, V. L. D.; SAMPAIO, M.J. A. M.; DOBEREINER, J. A fourth *Azospirillum* species from cereal roots. **An Acad. Bras. Cienc.**, v. 56, p. 365, 1984.

BALDANI, J.I.; BALDANI, V.L.D.; SELDIN, L; DOBEREINER, J. Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a root-associated nitrogen fixing bacterium. **Intl. J. Sys. Bact.**, v. 36, p. 86-93, 1986.

BALDANI, V. L. ; BALDANI, J. I. ; OLIVARES, F. L. ; DOBEREINER, J. Identification and ecology of *Herbaspirillum seropedicae* and *Burkholderia* spp. **Biology and Fertility of Soils**, v. 30, p. 485 – 489, 1992.

BALDANI, J.I.; OLIVARES, F.L.; DOBEREINER, J. Selection of *Herbaspirillum* spp. Strains associated with rice seedlings amended with ¹⁵N-labeled fertilizer. In: **International Symposium on Sustainable Agriculture for the tropics – The Role of Biological Nitrogen Fixation**. BODDEY, R. M.; DE RESENDE, A.S. Eds., EMBRAPA, P. 202-203, 1995.

BATUT, J.; DAVERAN-MINGOT, M. L.; DAVID, M.; JACOBS, J.; GARNERONE, A. M.; KAHN, D. *fixK*, a gene homologous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. **EMBO J.**, v. 8, p. 1279 - 1286, 1989.

BATUT, J.; BOISTARD, P. Oxygen control in *Rhizobium*. **Antonie Van Leeuwenhoek**, v. 66 (1-3), p. 129 - 150, 1994.

BODDEY, R. M.; OLIVEIRA, O. C.; URQUIAGA, S.; REIS, V. M.; OLIVARES, F. L.; BALDANI, V. L. D.; DOBEREINER, J. Biological nitrogen fixation associated with sugar cane and rice: contributions and prospects for improvement. **Plant and Soil**, v. 174, p. 195-209, 1995.

BRADFORD, M. M. A Rapid and Sensitive for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. **Analytical Biochemistry**, v. 72, p. 248-254, 1976.

BURGESS, B. K.; LOWE, D. J. Mechanism of Molybdenum Nitrogenase. **Chem. Rev.**, v. 96, p. 2983 – 3011, 1996.

COLLONO-ROMANO, S.; ARNOLD, W.; SCHLÜTER, A.; BOISTARD, P.; PÜHLER, A.; PRIEFER, U. B. An Fnr-like protein encoded in *Rhizobium leguminosarum biovar viciae* shows structural and functional homology to *Rhizobium meliloti fixK*. **Mol Gen. Genet.**, v. 223, p. 138 - 147, 1990.

DAVID, M.; DAVERAN, M. L.; BATUT, J.; DEDIEU, A.; DOMERGUE, O.; GHAI, J.; HERTIG, C.; BOISTARD, P.; KAHN, D. Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. **Cell**, v. 54, p. 671 - 683, 1988.

DIXON, R. The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in gamma-proteobacteria. **Arch. Microbiol.**, v. 169, p. 371-380, 1998.

DÖBEREINER, J. Isolation and identification of root associated diazotrophs. **Plant and soil**, v. 10, p. 207-212, 1988.

DÖBEREINER, J. Recent changes in concepts of plant bacteria interactions. Endophytic N₂ fixing bacteria. **Ciência e cultura**, v. 44, p. 310-313, 1992.

EARL, C.D.; RONSON, C. W., AUSUBEL, F. M. Genetic and Structural analysis of the *Rhizobium meliloti fixA*, *fixB*, *fixC* and *fixX* genes. **J. Bacteriol**, v. 169. p. 1127 - 1136, 1987.

EIGLIMEIER, K.; HONORÉ, N.; IUCHI, S.; LIN, E. C.; LIN, E. C. C. Molecular genetic analysis of FNR-dependent promoter. **Mol. Microbiol**, v. 3, p. 869 - 878, 1989.

FAOSTAT ; **Agricultural data**. Food and Agricultural Organization of the United Nations, Rome, 2001.

FISCHER, H. M. ; BRUDERER, T. ; HENNECKE, H. Essential and non-essential domains in the *Bradyrhizobium japonicum* NifA protein: identification of indispensable cysteine residues potentially involved in redox reactivity and/or metal binding. **Nucl Acids Res.**, v.16, p. 2207-2224, 1988.

FISCHER, H. Genetic regulation of nitrogen fixation in rhizobia. **Microbiol. Rev.**, v. 58, p. 352-386, 1994.

FOUSSARD, M.; GARNERONE, A. M.; NI, F. ; SOUPENE, E. ; BOISTARD, P. ; BATUT, J. Negative autorregulation of the *Rhizobium meliloti fixK* gene is indirect and requires a newly identified regulator, FixT. **Mol. Microbiol**, v. 25 (1), p. 25 – 37, 1997.

GARCIA-HORSMAN, J.A; Garry, E.; Shapleigh, J.P.; Alben, J.O.; Gennis, R.B. A novel cytochrome *c* oxidase from *Rhodobacter sphaeroides* that lacks *Cu_A*. **Biochemistry**. v. 33, p. 3113-3119, 1994.

HALBLEIB, C. M.; LUDDEN, P. W. Regulation of Biological Nitrogen Fixation. **American Society for Nutritional Sciences**, v. 130, p. 1081 – 1084, 2000.

KAHN, D. David, M. Domergue, O. Daveran, M.L. Ghai, J. Hirsch P.R. Batut, J. *Rhizobium meliloti fixGHI* sequence predicts involvement of a specific cation pump in symbiotic nitrogen fixation. **J. Bacteriol**. **171(2)**: 929–939, 1989.

KAHN, D. ; BATUT, J. ; DAVERAN, M. L. ; FOURMENT, J. Structure and regulation of the *fixNOQP* operon from *Rhizobium meliloti*, p. 474. In R. Palacios, J. Mora and W. E. Newton (ed.), *New horizons in nitrogen fixation*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.

KAMINSKI, P. A.; MANDON, K.; ARIGONI, F.; DESNOUES, N.; ELMERICH, C. Regulation of nitrogen fixation in *Azorhizobium caulinodans*: identification of a *fixK*-like gene, a positive regulator of *nifA*. **Mol Microbiol**, v. 5, p. 1983 – 1991, 1991.

KLASSEN, G.; PEDROSA, F. O.; SOUZA, E. M.; YATES, M. G.; RIGO, L. U. Sequencing and functional analysis of the *nifENXorf1orf2* gene cluster of *Herbaspirillum seropedicae*. **FEMS Microbiol. Lett.**, v. 181, p. 165-170, 1999.

LOPEZ, O. ; MORERA, C. ; MIRANDA-RIOS, J. ; GIRARD, L. ; ROMERO, D. ; SOBERÓN, M. Regulation of Gene Expression in Response to Oxygen in *Rhizobium etli*. Role of FnrN in *fixNOQP* Expression and in Symbiotic Nitrogen Fixation. **Amer Soc for Microbiol**, v. 183 (24), p. 6999 – 7006, 2001.

LOWE, D. J.; BURGESS, B. K. Mechanism of Molybdenum Nitrogenase. **American Chemical Society**, Rev. v. 96, p. 2983-3011, 1996.

MACHADO, I. M. P.; YATES, M.G.; MACHADO, H. B.; SOUZA, E. M.; PEDROSA, F. O. Cloning and sequence of the nitrogenase structural genes *nifHDK* of *Herbaspirillum seropedicae*. **Braz. J. Med. Biol. Res.**, v. 29, p. 1599-1602, 1996.

MANDON, K.; KAMINSKI, A.; ELMERICH, C. Functional Analysis of the *fixNOQP* Region of *Azorhizobium caulinodans*. **Journal of Bacteriology**, v. 176, p. 2560-2568, 1994.

MERRICK, M. J. Regulation and nitrogen fixation genes in free living and symbiotic bacteria. 1992 p. 835 – 876. In STACEY, G. ; BURRIS, R.H.; EVANS, H. J. (ed) **Biological nitrogen fixation**. Chapman & Hall, New York.

MERRICK, M. J. Organization and regulation of nitrogen fixation genes. In PALACIOS, R.; MORA, J.; NEWTON, W. E. (ed.) **New horizons in nitrogen fixation**. Kluwer Academic Publisher, Dordrecht, The Netherlands, p. 1 – 12, 1993.

MONTEIRO, R. A.; SOUZA, E. M.; FUNAYAMA, S.; YATES, M. G.; PEDROSA, F. O.; CHUBATSU, L. S. Expression and functional analysis of an N-truncated NifA protein of *Herbaspirillum seropedicae*. **FEBS Lett**, v. 447, p. 283-286, 1999.

MONTEIRO, R. A.; SOUZA, E. M.; YATES, M. G.; PEDROSA, F. O.; CHUBATSU, L. S. Fnr is involved in oxygen control of *Herbaspirillum seropedicae* N-truncated NifA protein activity in *Escherichia coli*. **Appl Environ Microbiol**. v. 69 (3), p 1527-1531, 2003.

MONTEIRO, R. A. Análise funcional dos domínios modulares da proteína NifA de *Herbaspirillum seropedicae*. Curitiba, 2001. Tese de Doutorado em Ciências – Bioquímica. Setor de Ciências Biológicas, Universidade Federal do Paraná.

NEWTON, W.E. Nitrogen fixation in perspective. In: PEDROSA, F.O.; HUNGRIA, M.; YATES, M.G. e NEWTON, W.E (Eds). **Nitrogen fixation: from molecules to crop productivity**. Proceedings of the 12th International Congress on Nitrogen Fixation, 12-17 de setembro de 1999, Foz do Iguaçu, Paraná, Brasil. P. 3-13, 2000.

OLIVARES, F. L.; DOS REIS, F. B.; REIS, V. M.; BALDANI, V. L.; BALDANI, J. I.; DOBEREINER, J. Infection of sugarcane roots by the endophytic diazotrophs *Herbaspirillum seropedicae* and *H. rubrisubalbicans*. In: **International Symposium on Sustainable Agriculture for the tropics – The Role of Biological Nitrogen Fixation**. BODDEY, R. M.; DE RESENDE, A. S. Eds., EMBRAPA , p.65-66, 1995.

OLIVARES, F. L. ; BALDANI, V. L. D. ; REIS,V. M. ; BALDANI, J. I.; DOBEREINER, J. Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stems and leaves predominantly of Gramineae. **Biol. Fertil. Soils**, v. 21, p. 197-200, 1996.

OLIVARES, F. L.; JAMES, E. K.; BALDANI, J. I.; DOBEREINER, J. Infection of mottled stripe disease susceptible and resistant varieties of sugarcane by the endophytic diazotroph *Herbaspirillum seropedicae*. **New Phytol.**, v. 135, p. 723-737, 1997.

PREISIG, O.; ANTHAMATTEN, D.; HENNECK, H. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for nitrogen-fixing endosymbiosis. **Proc Natl Acad Sci USA**, v. 90, p. 3309 – 3313, 1993.

POSTGATE, J. R. Biological nitrogen fixation: Fundamental. **Phill. Trans. R. Soc. Lond.** v. 296, p. 375-385, 1982.

RAEDER, K. R. Isolamento e caracterização do gene *fnr* de *Herbaspirillum seropedicae*, Curitiba 2003. Tese de Mestrado em Ciências-Bioquímica, Setor de Ciências Biológicas, Universidade Federal do Paraná.

REGO, F. G. M. Sequenciamento do gene *nifB* de *Herbaspirillum seropedicae* e caracterização da sua região promotora. Curitiba, 1997. 143p. Tese de mestrado em Ciências-Bioquímica. Setor de Ciências biológicas, Universidade Federal do Paraná.

SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. **Molecular cloning a laboratory manual**. 2 ed. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 1989.

SCHLESINGER, W. H. **Biogeochemistry: An analysis of global change**, 2ª edição. Academic Press, San Diego, CA. 1997.

SHARMA, V.; PUUSTINEN, A.; WIKSTRÖM, M.; LAAKKONEN, L. Sequence Analysis of the *cbb₃* Oxidases and an Atomic Model for the *Rhodobacter sphaeroides* Enzyme. **Biochemistry**, v.45, p. 5754-5765, 2006.

SIMPSON, F. B.; BURRIS, R. H. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. **Science**, v. 224, p. 1095-1096, 1984.

SOUZA, E. M.; FUNAYAMA, S.; RIGO, L.U.; YATES, M.G.; PEDROSA, F.O. Sequence and strutral organization of a *nifA*-like gene and part of a *nifB*-like gene of *Herbaspirillum seropedicae* strain Z78. **J. Gen. Microb.**, Reading, v. 137, p. 1511-1522, 1991a.

SOUZA, E. M.; FUNAYAMA, S.; RIGO, L. U.; PEDROSA, F. O. Cloning and characterization of the *nifA* gene from *Herbaspirillum seropedicae* strain Z78. **Can. J. Microbiol.**, v.37, p. 425-429, 1991b.

SOUZA, E. M.; PEDROSA, F. O.; MACHADO, H. B.; DRUMMOND, M.; YATES, M. G. The N-terminus of the NifA protein of *H. seropedicae* is probably involved in sensing of ammonia. In: **Nitrogen Fixation: Fundamentals and Applications**. TIKHONOVICH, I. A.; NIKOLAI, A. P.; VASSILY, I.R.; WILLIAM, E. N. (eds). X International Congress in Nitrogen Fixation. St Petersburg, p. 260, 1995.

SOUZA, E. M.; PEDROSA, F. O.; DRUMMOND, M.; RIGO, L. U.; YATES, M. G. Control of *Herbaspirillum seropedicae* NifA activity by ammonium ions and oxygen. *J. Bacteriol.* V.181, p. 681-684, 1999.

TAIZ, L.; ZEIGER, E. **Fisiologia Vegetal**. 3º Edição. Editora ArtMed, 2004.

ZUFFEREY, R.; PREISG, O.; HENNECKE, H.; MEYER, T. Assembly and Function of the Cytochrome *cbb₃* Oxidase Subunits in *Bradyrhizobium japonicum*. **The Journal of Biological Chemistry**, v. 271, p.9114-9119, 1996.

APÊNDICE 1 – SEQUÊNCIA DO OPERON *fixISNOQPGH* DE *H. seropedicae* ANALISADA

| | | |
|------|---|------|
| 1 | atg gac cag cgt ggt ccg cgc tac acg tcc tat ccc act gca gac | 45 |
| 46 | cgt ttc acc agt gac ttt gcc gtc acc gac tac ctc cat gcg gtg | 90 |
| 91 | tcc gac cgt cgc agc atg agc gcc tgg cgc gcc ttg tcg ctg tac | 135 |
| 136 | ctg cac att ccg ttc tgc gac acc atc tgc tat tac tgc gcc tgc | 180 |
| 181 | aac aag atc gtc acc aag aac cgc gcc aag gcc gcg ctc tac ctg | 225 |
| 226 | agc tac ctc aag cgc gag atc agc atg cag ggc tcg ctc ttt tcg | 270 |
| 271 | gga atg aac cag gtc gag caa ctg cac ttc ggc ggc ggc acc ccg | 315 |
| 316 | acc tac ctt tcg gat gag cag atg tcg gac ctg atg gac cac atc | 360 |
| 361 | cgc cac tgc ttc acg ctc gca cct gac cac gtc ggc gaa tac tcc | 405 |
| 406 | atc gag atc gat ccc cgc acg gtg tcg gtg gaa cgc gtc cac aag | 450 |
| 451 | ctg cgc cag cag ggc ttc aac cgc atc agc ctg ggc gtg cag gat | 495 |
| 496 | ttc gat ccc gag gtg cag ctg gcc gtg aac cgt gtc cag tcg gaa | 540 |
| 541 | gag cag acg ctg gag atc atc cgc gcc gcc cgc gat gcc ggt ttc | 585 |
| 586 | cgc tcg gtc agc atc gac ttg atc tat ggc ttg ccc aag cag aac | 630 |
| 631 | gtg atg tcc atg tcg cgc acg ctg gcc aag gtg atc gcc gcc agc | 675 |
| 676 | ccg gac cgc atc gcc gtc tac aac tat gcg cac atg ccg cag ctc | 720 |
| 721 | ttc aag acc cag cgc cag atc aag gaa gaa gac ctg ccc agc gcc | 765 |
| 766 | gac agc aag ctc gac atg ctc tcg ctg tgc atc cgc cag ctg acc | 810 |
| 811 | tcg gcg ggc tat gtc tat atc ggc atg gat cac ttc gcc aag ccc | 855 |
| 856 | act gac gac ctg gcc att gcc cag cag cag ggt cgt ctg cac cgc | 900 |
| 901 | aac ttc cag ggt tac tcc acc cat tcg gag acc gac ctc gtg gcc | 945 |
| 946 | tgc ggc gta tcg gcc atc agc gcg gtg ggc ggc agc tac agc cag | 990 |
| 991 | aac gaa aag acc ctc gac gac tac tac gcc cgc ctg gaa aat tcc | 1035 |
| 1036 | acg ctg ccc atc gcc cgt ggc atc cag ctg ggc atg gat gat gtc | 1080 |
| 1081 | ttg cgc cgc ctg atc atc cag cgc ctg atg tgc aat ttc gag ttg | 1125 |
| 1126 | tcg atc aac tcg ctg gaa atc gcc tat ccc atc gtc ttc cgc gaa | 1170 |
| 1171 | tac ttc gcc agc gag atg gaa aag ctc aag cag ctg gaa gaa gac | 1215 |
| 1216 | ggc ctc atc aag atc gaa ccg gaa tgg atc acc gtc gag ccc aag | 1260 |
| 1261 | ggc cgg ctg ctg atc cgc aac atc tgc atg gtc ttc gat cgc tac | 1305 |

| | | |
|------|--|-------------|
| 1306 | ctg aac cag gag cgc gag aag gcc gcc agc cag aac aag gat gcc | 1350 |
| 1351 | ccg cag cga tat tgc cag acg gtc tga <u>gcg gga gcg</u> gga <u>gtg aat</u> Val Asn | 1395 465 |
| 1396 | ctg ctg ccg atc ttc ctg gta ggc ctg atg ggg agc gtc cat tgc | 1440 |
| 466 | Leu Leu Pro Ile Phe Leu Val Gly Leu Met Gly Ser Val His Cys | 480 |
| 1441 | atc ggg atg tgt ggc ggc atc gtc ggc gcg ctc agt tgc gcg gcg | 1485 |
| 481 | Ile Gly Met Cys Gly Gly Ile Val Gly Ala Leu Ser Ser Ala Ala | 495 |
| 1486 | ccc gtt cgt ccg acg ccc acg ccg gct gcg ccc ctg ctg tgc gtg | 1530 |
| 496 | Pro Val Arg Pro Thr Pro Thr Pro Ala Ala Pro Leu Leu Ser Val | 510 |
| 1531 | gcc gcg ccc ggt cgc aat gac ggc gcg cga ctg ctg cgc atc att | 1575 |
| 511 | Ala Ala Pro Gly Arg Asn Asp Gly Ala Arg Leu Leu Arg Ile Ile | 525 |
| 1576 | ccc att cat ccc gcc ggt ggc gtc cag gtc cag gcc atg gcc cag | 1620 |
| 526 | Pro Ile His Pro Ala Gly Gly Val Gln Val Gln Ala Met Ala Gln | 540 |
| 1621 | gat ctg gtg cgg gtg ctg tgc tac aac ctg ggc cgc ctg tcc agc | 1665 |
| 541 | Asp Leu Val Arg Val Leu Cys Tyr Asn Leu Gly Arg Leu Ser Ser | 555 |
| 1666 | tat gcg ctg gcc ggt gca ttg gcc ggc ggt atc gcg gca ggg ttg | 1710 |
| 556 | Tyr Ala Leu Ala Gly Ala Leu Ala Gly Gly Ile Ala Ala Gly Leu | 570 |
| 1711 | ctg cgc ggt gcc gat gtg ctc ggc tgg ctg gcg ccc gcg cag cgg | 1755 |
| 571 | Leu Arg Gly Ala Asp Val Leu Gly Trp Leu Ala Pro Ala Gln Arg | 585 |
| 1756 | gtg gct tac ctg atc acc aat atc gtg ctg gtc ctg ctg ggt ctg | 1800 |
| 586 | Val Ala Tyr Leu Ile Thr Asn Ile Val Leu Val Leu Leu Gly Leu | 600 |
| 1801 | tac ctg acc caa tgg tgg ccg gcg ctg gcg cgt ctg gag cag tgg | 1845 |
| 601 | Tyr Leu Thr Gln Trp Trp Pro Ala Leu Ala Arg Leu Glu Gln Trp | 615 |
| 1846 | ggc agc gcc ttg tgg gtg cgt gta cgc ccg ctg gcc gcg cgc ctg | 1890 |
| 616 | Gly Ser Ala Leu Trp Val Arg Val Arg Pro Leu Ala Ala Arg Leu | 630 |
| 1891 | gtg ccg gtg gat acg ccg gcc aag gcc ttg ctg ctg ggc agc ctc | 1935 |
| 631 | Val Pro Val Asp Thr Pro Ala Lys Ala Leu Leu Leu Gly Ser Leu | 645 |
| 1936 | tgg ggt tgg ctg ccc tgc ggg atg gtc tac agc gcc ttg ttg acg | 1980 |
| 646 | Trp Gly Trp Leu Pro Cys Gly Met Val Tyr Ser Ala Leu Leu Thr | 660 |
| 1981 | gcc ttg atg gcc ggc agc gcc atg cag ggg gca ttg acc atg ctg | 2025 |
| 661 | Ala Leu Met Ala Gly Ser Ala Met Gln Gly Ala Leu Thr Met Leu | 675 |
| 2026 | gcc ttt ggc gcc ggc acc ttg cca gtg ttg ctg gcg gct ggc ctg | 2070 |
| 676 | Ala Phe Gly Ala Gly Thr Leu Pro Val Leu Leu Ala Ala Gly Leu | 690 |
| 2071 | tcc ggt gcg cgc ctg cgc cag ctg gcg cag cgt ccg gcg gtg cgt | 2115 |
| 691 | Ser Gly Ala Arg Leu Arg Gln Leu Ala Gln Arg Pro Ala Val Arg | 705 |
| 2116 | ctg gcg gcc ggc gtg gtc gta ctg gct ttc gga ata atg ggt ttg | 2160 |
| 706 | Leu Ala Ala Gly Val Val Val Leu Ala Phe Gly Ile Met Gly Leu | 720 |
| 2161 | ctg cgt ggc gcc gag atc ggc ggc atg ggc att gca cgc ggc tgg | 2205 |
| 721 | Leu Arg Gly Ala Glu Ile Gly Gly Met Gly Ile Ala Arg Gly Trp | 735 |
| 2206 | atc gat gtg ttc tgc atc agt ccg gcg cat ggg gga tga cga tga | 2250 |
| 736 | Ile Asp Val Phe Cys Ile Ser Pro Ala His Gly Gly End | 750 |
| 2251 | gca gtg aag gga gcg ggg acg gga gcg gcg <u>agg gga</u> cac tgt gtt | 2295 |

| | | | | | | | | | | | | | | | | |
|------|-----|-----|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|
| 2296 | tcc | att | <u>gcg gcc</u> | agc | gtg | ccg | gcc | gga | cag | gcg | tgg | acg | ctg | gag | 2342 | |
| | | | | | Val | Pro | Ala | Gly | Gln | Ala | Trp | Thr | Leu | Glu | 779 | |
| 2343 | atc | ggc | ggg | gcg | cgy | cgt | gag | ctg | tgc | tgc | gtc | ggt | tgc | cag | agc | 2387 |
| 780 | Ile | Gly | Gly | Ala | Arg | Arg | Glu | Leu | Cys | Cys | Val | Gly | Cys | Gln | Ser | 794 |
| 2388 | gtg | gcg | cga | ctg | atc | gtg | gac | agc | ggc | tgc | gaa | gac | ttc | tac | ctg | 2432 |
| 795 | Val | Ala | Arg | Leu | Ile | Val | Asp | Ser | Gly | Cys | Glu | Asp | Phe | Tyr | Leu | 809 |
| 2433 | cgc | cgc | acg | gcg | cct | tcg | gcc | cgg | gtc | gat | ccc | gaa | caa | ctg | ctc | 2477 |
| 810 | Arg | Arg | Thr | Ala | Pro | Ser | Ala | Arg | Val | Asp | Pro | Glu | Gln | Leu | Leu | 824 |
| 2478 | ccg | ccc | gaa | ctg | gcc | ttg | ctg | gag | cag | ccg | cag | cca | gcg | ccg | gat | 2522 |
| 825 | Pro | Pro | Glu | Leu | Ala | Leu | Leu | Glu | Gln | Pro | Gln | Pro | Ala | Pro | Asp | 839 |
| 2523 | cag | gcc | cat | gcc | gcc | gac | gcc | agc | gcg | gag | ctg | gtc | ctg | tcc | atc | 2567 |
| 840 | Gln | Ala | His | Ala | Ala | Asp | Ala | Ser | Ala | Glu | Leu | Val | Leu | Ser | Ile | 854 |
| 2568 | gat | ggc | ttg | cgc | tgc | tcg | gcc | tgc | gtc | tgg | ctg | atc | gaa | aaa | tac | 2612 |
| 855 | Asp | Gly | Leu | Arg | Cys | Ser | Ala | Cys | Val | Trp | Leu | Ile | Glu | Lys | Tyr | 869 |
| 2613 | ctg | gcg | cgc | ctg | ccc | ggc | atg | cag | atg | gcc | gag | atg | aac | gtg | gct | 2657 |
| 870 | Leu | Ala | Arg | Leu | Pro | Gly | Met | Gln | Met | Ala | Glu | Met | Asn | Val | Ala | 884 |
| 2658 | acc | tca | cgg | ctg | cac | ctg | cgc | cgc | gat | ccc | gcc | ttg | tgt | tcc | acc | 2702 |
| 885 | Thr | Ser | Arg | Leu | His | Leu | Arg | Arg | Asp | Pro | Ala | Leu | Cys | Ser | Thr | 899 |
| 2703 | gcc | acc | atc | ctg | cgc | ggc | ctg | cgc | ggg | ctg | ggc | tat | acc | gcc | tat | 2747 |
| 900 | Ala | Thr | Ile | Leu | Arg | Gly | Leu | Arg | Gly | Leu | Gly | Tyr | Thr | Ala | Tyr | 914 |
| 2748 | ccc | ttc | gat | ccc | ctg | cgc | cag | ggt | gaa | cag | gcg | cgc | cgt | gcc | tcg | 2792 |
| 915 | Pro | Phe | Asp | Pro | Leu | Arg | Gln | Gly | Glu | Gln | Ala | Arg | Arg | Ala | Ser | 929 |
| 2793 | cgc | cgc | ctg | ttc | cgc | cag | ttg | ttc | atc | gcc | ggc | ctg | tcg | atg | atg | 2837 |
| 930 | Arg | Arg | Leu | Phe | Arg | Gln | Leu | Phe | Ile | Ala | Gly | Leu | Ser | Met | Met | 944 |
| 2838 | cag | gtg | atg | atg | tat | gcc | gtg | ccg | gtc | tac | atg | acc | cat | gag | ggc | 2882 |
| 945 | Gln | Val | Met | Met | Tyr | Ala | Val | Pro | Val | Tyr | Met | Thr | His | Glu | Gly | 959 |
| 2883 | atc | gat | ccc | gac | atg | atg | tcg | ctc | atg | cgc | tgg | gcc | agc | ctg | ttc | 2927 |
| 960 | Ile | Asp | Pro | Asp | Met | Met | Ser | Leu | Met | Arg | Trp | Ala | Ser | Leu | Phe | 974 |
| 2928 | ctg | acc | att | cct | gcg | gtg | ttc | tat | tcg | gcc | ttg | cca | ttc | ttc | acc | 2972 |
| 975 | Leu | Thr | Ile | Pro | Ala | Val | Phe | Tyr | Ser | Ala | Leu | Pro | Phe | Phe | Thr | 989 |
| 2973 | ggc | gcc | tgg | gcc | gga | ctg | cgc | gcc | cgt | gca | ccg | ggg | atg | gac | ctg | 3017 |
| 990 | Gly | Ala | Trp | Ala | Gly | Leu | Arg | Ala | Arg | Ala | Pro | Gly | Met | Asp | Leu | 1004 |
| 3018 | ccg | gtg | gcc | atc | ggc | atc | gca | gca | gcc | ttc | ctg | gcc | agc | gcc | atc | 3062 |
| 1005 | Pro | Val | Ala | Ile | Gly | Ile | Ala | Ala | Ala | Phe | Leu | Ala | Ser | Ala | Ile | 1019 |
| 3063 | gcc | acc | tgg | cgc | ggc | gag | ggc | gag | atc | tgg | ttc | gac | agc | gtg | agc | 3107 |
| 1020 | Ala | Thr | Trp | Arg | Gly | Glu | Gly | Glu | Ile | Trp | Phe | Asp | Ser | Val | Ser | 1034 |
| 3108 | atg | ttc | atc | ttc | ctg | ctg | ctg | ggc | agt | cgc | tac | ctg | gag | gcg | gcc | 3152 |
| 1035 | Met | Phe | Ile | Phe | Leu | Leu | Leu | Gly | Ser | Arg | Tyr | Leu | Glu | Ala | Ala | 1049 |
| 3153 | gcc | cgc | cgc | aaa | tcc | gcg | tcg | gcg | ctg | gag | cgc | atg | cag | cag | gcc | 3197 |
| 1050 | Ala | Arg | Arg | Lys | Ser | Ala | Ser | Ala | Leu | Glu | Arg | Met | Gln | Gln | Ala | 1064 |
| 3198 | ttc | ccg | gcc | tcg | gcg | ctg | ctg | gcg | cca | ggc | tat | ccg | gag | cag | cgc | 3242 |

| | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 1065 | Phe | Pro | Ala | Ser | Ala | Leu | Leu | Ala | Pro | Gly | Tyr | Pro | Glu | Gln | Arg | 1079 |
| 3243 | gag | acc | acc | ctg | gtg | gtc | gca | gcg | caa | ttg | cgc | gag | ggt | gat | gtg | 3287 |
| 1080 | Glu | Thr | Thr | Leu | Val | Val | Ala | Ala | Gln | Leu | Arg | Glu | Gly | Asp | Val | 1094 |
| 3288 | atc | ctg | gcg | cgt | ccg | ggc | gat | acc | atc | gcg | gct | gac | gcc | tgc | ctg | 3332 |
| 1095 | Ile | Leu | Ala | Arg | Pro | Gly | Asp | Thr | Ile | Ala | Ala | Asp | Ala | Cys | Leu | 1109 |
| 3333 | ctg | gag | ggc | gcg | ggc | gag | ttc | gac | ctg | gcc | ctg | ctg | tgc | ggc | gaa | 3377 |
| 1110 | Leu | Glu | Gly | Ala | Gly | Glu | Phe | Asp | Leu | Ala | Leu | Leu | Ser | Gly | Glu | 1124 |
| 3378 | agc | cgt | ccc | cag | tct | ttc | cag | atc | ggc | cag | gaa | gcg | ccc | ggc | ggc | 3422 |
| 1125 | Ser | Arg | Pro | Gln | Ser | Phe | Gln | Ile | Gly | Gln | Glu | Ala | Pro | Gly | Gly | 1139 |
| 3423 | gcg | gtc | aac | ctg | agc | cag | ccg | gtg | ttc | ctg | cgg | gtg | gtg | cgc | gcc | 3467 |
| 1140 | Ala | Val | Asn | Leu | Ser | Gln | Pro | Val | Phe | Leu | Arg | Val | Val | Arg | Ala | 1154 |
| 3468 | acc | ggc | gaa | agt | acg | ctg | gcc | gcg | ctg | acc | cgg | ctg | gcc | gag | cag | 3512 |
| 1155 | Thr | Gly | Glu | Ser | Thr | Leu | Ala | Ala | Leu | Thr | Arg | Leu | Ala | Glu | Gln | 1169 |
| 3513 | gcc | gga | cag | ggc | aag | cct | gcg | ctg | gcc | cag | tgg | gct | gac | gtc | gtg | 3557 |
| 1170 | Ala | Gly | Gln | Gly | Lys | Pro | Ala | Leu | Ala | Gln | Trp | Ala | Asp | Val | Val | 1184 |
| 3558 | gct | tgc | cgc | ttc | gta | gtg | gcg | ctg | ctg | ctg | ctg | gct | ggg | gcg | act | 3602 |
| 1185 | Ala | Ser | Arg | Phe | Val | Val | Ala | Leu | Leu | Leu | Leu | Ala | Gly | Ala | Thr | 1199 |
| 3603 | ttc | ctg | gcc | tgg | caa | ttg | atc | gat | ccg | acg | cgt | gcc | tgg | tcc | acc | 3647 |
| 1200 | Phe | Leu | Ala | Trp | Gln | Leu | Ile | Asp | Pro | Thr | Arg | Ala | Trp | Ser | Thr | 1214 |
| 3648 | gcc | att | gcg | gtg | ctg | gtg | gtg | tct | tgt | ccc | tgc | gcc | ctg | tgc | ctg | 3692 |
| 1215 | Ala | Ile | Ala | Val | Leu | Val | Val | Ser | Cys | Pro | Cys | Ala | Leu | Ser | Leu | 1229 |
| 3693 | gcg | aca | ccc | tgc | gca | ttg | gcg | gcc | gcc | act | gat | cgt | ctg | ctg | cgg | 3737 |
| 1230 | Ala | Thr | Pro | Ser | Ala | Leu | Ala | Ala | Ala | Thr | Asp | Arg | Leu | Leu | Arg | 1244 |
| 3738 | gcc | ggt | acg | ctg | atc | gtg | cgt | ggc | aat | gtg | ctg | gaa | acc | ctg | caa | 3782 |
| 1245 | Ala | Gly | Thr | Leu | Ile | Val | Arg | Gly | Asn | Val | Leu | Glu | Thr | Leu | Gln | 1259 |
| 3783 | cgc | gcc | gac | acc | att | gtc | ttc | gac | aag | acc | ggc | acg | ctc | acc | cag | 3827 |
| 1260 | Arg | Ala | Asp | Thr | Ile | Val | Phe | Asp | Lys | Thr | Gly | Thr | Leu | Thr | Gln | 1274 |
| 3828 | gga | cgt | ccg | cag | ttg | acc | gcc | ttg | tgg | tgc | gat | gtg | gac | tgc | cag | 3872 |
| 1275 | Gly | Arg | Pro | Gln | Leu | Thr | Ala | Leu | Trp | Ser | Asp | Val | Asp | Cys | Gln | 1289 |
| 3873 | cgc | tgc | ctg | gcg | atc | gcc | gct | gcc | atg | gag | cgc | ggc | agc | ctg | cat | 3917 |
| 1290 | Arg | Ser | Leu | Ala | Ile | Ala | Ala | Ala | Met | Glu | Arg | Gly | Ser | Leu | His | 1304 |
| 3918 | ccg | ctg | gcc | aag | gcg | ctg | gtg | gag | gag | ggc | gcg | cgc | cgg | cag | gtg | 3962 |
| 1305 | Pro | Leu | Ala | Lys | Ala | Leu | Val | Glu | Glu | Gly | Ala | Arg | Arg | Gln | Val | 1319 |
| 3963 | gac | gcc | ctt | gag | gtg | gaa | caa | ctg | act | tgc | gtg | acg | ggg | gcg | ggc | 4007 |
| 1320 | Asp | Ala | Leu | Glu | Val | Glu | Gln | Leu | Thr | Ser | Val | Thr | Gly | Ala | Gly | 1334 |
| 4008 | atg | cag | ggc | atg | atc | gag | ggc | gac | ctc | tat | cgc | atc | ggc | tgc | cgc | 4052 |
| 1335 | Met | Gln | Gly | Met | Ile | Glu | Gly | Asp | Leu | Tyr | Arg | Ile | Gly | Ser | Arg | 1349 |
| 4053 | agt | ttc | gtc | gct | ggg | ctg | gcg | ggc | act | gca | ttg | cca | ccg | gtc | ctg | 4097 |
| 1350 | Ser | Phe | Val | Ala | Gly | Leu | Ala | Gly | Thr | Ala | Leu | Pro | Pro | Val | Leu | 1364 |
| 4098 | ttg | cag | tgc | gct | gcc | agt | ggg | gcg | ggc | agc | gtc | tat | ctg | ggt | agc | 4142 |
| 1365 | Leu | Gln | Ser | Ala | Ala | Ser | Gly | Ala | Gly | Ser | Val | Tyr | Leu | Gly | Ser | 1379 |
| 4143 | agc | gat | ggt | tgg | ctg | gcg | cgc | ttc | gac | ctg | gcc | gat | gcc | ttg | cgc | 4187 |


```

1380 Ser Asp Gly Trp Leu Ala Arg Phe Asp Leu Ala Asp Ala Leu Arg 1394
4188 acg gat gcg gtc gcg acc gtg gcg gct ttc cgc gca ctg ggc ctg 4232
1395 Thr Asp Ala Val Ala Thr Val Ala Ala Phe Arg Ala Leu Gly Leu 1409
4233 cgc acc atc ctg ctc agc ggc gac cag ccc gag gtc tgt gcc gag 4277
1410 Arg Thr Ile Leu Leu Ser Gly Asp Gln Pro Glu Val Cys Ala Glu 1424
4278 gtg gcc gcc gca gcc ggc atc agc gag gtg cag gca gcc tgc acg 4322
1425 Val Ala Ala Ala Ala Gly Ile Ser Glu Val Gln Ala Ala Cys Thr 1439
4323 cca cag ggc aag ctg gac tat gtg cgc aat ctg cag cgg gac ggg 4367
1440 Pro Gln Gly Lys Leu Asp Tyr Val Arg Asn Leu Gln Arg Asp Gly 1454
4368 gcg gtg gtg gcc atc gtt ggc gat ggc atc aac gac gcg gcc atg 4412
1455 Ala Val Val Ala Ile Val Gly Asp Gly Ile Asn Asp Ala Ala Met 1469
4413 ctg cgc gcc ggc gat gtc tcc ttc gcc atg ggc aag ggc gcg gcg 4457
1470 Leu Arg Ala Gly Asp Val Ser Phe Ala Met Gly Lys Gly Ala Ala 1484
4458 ctg gcg cag gtc agt gcc gat gcc gtc atc atg agt gac cgc ctg 4502
1485 Leu Ala Gln Val Ser Ala Asp Ala Val Ile Met Ser Asp Arg Leu 1499
4503 cag gcc gtg gcc gat tgc gcg gcg atg gcg cag cgg acc atg cac 4547
1500 Gln Ala Val Ala Asp Cys Ala Ala Met Ala Gln Arg Thr Met His 1514
4548 atc gtg cgc cag aac ctg gtg tgg gcg gcg acc tac aac ttc ctg 4592
1515 Ile Val Arg Gln Asn Leu Val Trp Ala Ala Thr Tyr Asn Phe Leu 1529
4593 gcc att ccg gca gca gcc ttc ggc ctg ctc gat ccc tgg atg tgc 4637
1530 Ala Ile Pro Ala Ala Ala Phe Gly Leu Leu Asp Pro Trp Met Ser 1544
4638 gcg gtg ggc atg tgc ctc agt tgc ctg ctg gtg gtg ggc aat gcc 4682
1545 Ala Val Gly Met Ser Leu Ser Ser Leu Leu Val Val Gly Asn Ala 1559
4683 ttg cgc ctg agc cgt cgg cag gtg cgt tct gct gca gaa gca cag 4727
1560 Leu Arg Leu Ser Arg Arg Gln Val Arg Ser Ala Ala Glu Ala Gln 1574
4728 ccg gcg gcg gta cgt ccc gcg ctg gcg gga gcg gcc tga tg gaa 4771
1575 Pro Ala Ala Val Arg Pro Ala Leu Ala Gly Ala Ala End Glu 1590
4772 gcg ctc tat ctc ctc att ccc ctg agc acc ctg ctg gtc ttc ctg 4816
1591 Ala Leu Tyr Leu Leu Ile Pro Leu Ser Thr Leu Leu Val Phe Leu 1605
4817 gcg atc tgg gtg ttc ttc cgc gcc tcc gac agc ggc cag ttc gat 4861
1606 Ala Ile Trp Val Phe Phe Arg Ala Ser Asp Ser Gly Gln Phe Asp 1620
4862 gac ctc gaa ggg cct gcc atg cgc atc ctc cac gac gat gat gcc 4906
1621 Asp Leu Glu Gly Pro Ala Met Arg Ile Leu His Asp Asp Asp Ala 1635
4907 gcc gtc ccg caa gag cca tcc ggc ggc gcg tga tt gtg tgc act 4950
1636 Ala Val Pro Gln Glu Pro Ser Gly Gly Ala End
4951 gcg gtt gac gca gac cgt tgt ttc ccc gtc ttt gtt ccc gcg ctt 4995
4996 ctc ccc att ttg ata cac gtc aag gct gcc gga agc tgc gcc ccg 5040
5041 taa gct cgc att gac cag tac tca aca tag tgg gga gag tttcgtg 5086
Val 1695
5087 agc aaa gaa aat agc tac aac tac acg gtt gtg cgc caa ttc acc 5131
1696 Ser Lys Glu Asn Ser Tyr Asn Tyr Thr Val Val Arg Gln Phe Thr 1710

```

| | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 5132 | gtg | gcg | acc | atc | ctg | tgg | ggc | gta | gtc | ggg | atg | ttg | gtc | ggc | gtg | 5176 |
| 1711 | Val | Ala | Thr | Ile | Leu | Trp | Gly | Val | Val | Gly | Met | Leu | Val | Gly | Val | 1725 |
| 5177 | atc | atc | gcc | gcc | caa | ctg | gca | tgg | cct | gaa | ctg | aac | atg | ggg | gta | 5221 |
| 1726 | Ile | Ile | Ala | Ala | Gln | Leu | Ala | Trp | Pro | Glu | Leu | Asn | Met | Gly | Val | 1740 |
| 5222 | ccc | tgg | ctg | agc | ttc | ggc | cgg | ctg | agg | ccc | ctg | cat | acc | aac | gcg | 5266 |
| 1741 | Pro | Trp | Leu | Ser | Phe | Gly | Arg | Leu | Arg | Pro | Leu | His | Thr | Asn | Ala | 1755 |
| 5267 | gtg | atc | ttc | gcc | ttc | ggc | ggc | tgt | gcg | ctg | atg | gcc | act | tca | tat | 5311 |
| 1756 | Val | Ile | Phe | Ala | Phe | Gly | Gly | Cys | Ala | Leu | Met | Ala | Thr | Ser | Tyr | 1770 |
| 5312 | tac | gtc | gtc | cag | cgc | acc | tgc | cag | gtg | cgc | ctg | ttc | tcc | gac | ttc | 5356 |
| 1771 | Tyr | Val | Val | Gln | Arg | Thr | Cys | Gln | Val | Arg | Leu | Phe | Ser | Asp | Phe | 1785 |
| 5357 | ctg | gcc | gcc | ttc | acc | ttc | tgg | ggg | tgg | cag | ctg | gtg | atc | gtc | ggc | 5401 |
| 1786 | Leu | Ala | Ala | Phe | Thr | Phe | Trp | Gly | Trp | Gln | Leu | Val | Ile | Val | Gly | 1800 |
| 5402 | gcc | gcc | atc | tcg | ctg | ccg | ctc | ggg | ctg | acc | cgt | ggc | aag | gaa | tac | 5446 |
| 1801 | Ala | Ala | Ile | Ser | Leu | Pro | Leu | Gly | Leu | Thr | Arg | Gly | Lys | Glu | Tyr | 1815 |
| 5447 | gcc | gaa | ctg | gaa | tgg | ccc | ctg | acc | atc | ctg | atc | gcc | gtc | gtc | tgg | 5491 |
| 1816 | Ala | Glu | Leu | Glu | Trp | Pro | Leu | Thr | Ile | Leu | Ile | Ala | Val | Val | Trp | 1830 |
| 5492 | gtg | gcc | tac | gcg | gtg | gtg | ttc | ttc | ggc | acc | ctg | atc | aag | cgc | aag | 5536 |
| 1831 | Val | Ala | Tyr | Ala | Val | Val | Phe | Phe | Gly | Thr | Leu | Ile | Lys | Arg | Lys | 1845 |
| 5537 | gtg | cag | cac | atc | tac | gtg | gcc | aac | tgg | ttc | tac | ggc | gcc | ttc | att | 5581 |
| 1846 | Val | Gln | His | Ile | Tyr | Val | Ala | Asn | Trp | Phe | Tyr | Gly | Ala | Phe | Ile | 1860 |
| 5582 | atc | gcc | gtg | gcc | atc | ctg | cac | atc | gtc | aac | ggc | atg | acc | atg | ccg | 5626 |
| 1861 | Ile | Ala | Val | Ala | Ile | Leu | His | Ile | Val | Asn | Gly | Met | Thr | Met | Pro | 1875 |
| 5627 | gcc | acc | ctg | acc | aag | tcg | tac | tcc | atg | tac | agc | ggc | gcg | cag | gat | 5671 |
| 1876 | Ala | Thr | Leu | Thr | Lys | Ser | Tyr | Ser | Met | Tyr | Ser | Gly | Ala | Gln | Asp | 1890 |
| 5672 | gcg | atg | atc | cag | tgg | tgg | tac | ggc | cac | aat | gcc | gtg | ggc | ttc | ttc | 5716 |
| 1891 | Ala | Met | Ile | Gln | Trp | Trp | Tyr | Gly | His | Asn | Ala | Val | Gly | Phe | Phe | 1905 |
| 5717 | ctg | acc | gcc | ggc | ttc | ctg | ggc | atg | atg | tac | tac | ttc | att | ccc | aag | 5761 |
| 1906 | Leu | Thr | Ala | Gly | Phe | Leu | Gly | Met | Met | Tyr | Tyr | Phe | Ile | Pro | Lys | 1920 |
| 5762 | cag | gtc | aac | cgt | ccg | gtg | tat | tcc | tac | cgt | ctg | tcg | atc | gtg | cac | 5806 |
| 1921 | Gln | Val | Asn | Arg | Pro | Val | Tyr | Ser | Tyr | Arg | Leu | Ser | Ile | Val | His | 1935 |
| 5807 | ttc | tgg | gcg | ctg | atc | ttc | acc | tac | atg | tgg | gcc | ggc | ccg | cac | cat | 5851 |
| 1936 | Phe | Trp | Ala | Leu | Ile | Phe | Thr | Tyr | Met | Trp | Ala | Gly | Pro | His | His | 1950 |
| 5852 | ctg | cat | tac | acc | gcg | ctg | cct | gac | tgg | acc | cag | tcg | ctg | ggc | atg | 5896 |
| 1951 | Leu | His | Tyr | Thr | Ala | Leu | Pro | Asp | Trp | Thr | Gln | Ser | Leu | Gly | Met | 1965 |
| 5897 | gtg | ttc | tcg | ctg | atc | ctg | ctg | gcg | ccg | tcc | tgg | ggg | ggc | atg | atc | 5941 |
| 1966 | Val | Phe | Ser | Leu | Ile | Leu | Leu | Ala | Pro | Ser | Trp | Gly | Gly | Met | Ile | 1980 |
| 5942 | aac | ggc | atg | atg | acc | ctg | tcg | ggc | gcc | tgg | cac | cag | ctg | cgc | acc | 5986 |
| 1981 | Asn | Gly | Met | Met | Thr | Leu | Ser | Gly | Ala | Trp | His | Gln | Leu | Arg | Thr | 1995 |
| 5987 | gat | ccc | atc | ctg | aag | ttc | ctg | gtg | gtg | tcg | ctg | tcc | ttc | tac | ggc | 6031 |
| 1996 | Asp | Pro | Ile | Leu | Lys | Phe | Leu | Val | Val | Ser | Leu | Ser | Phe | Tyr | Gly | 2010 |
| 6032 | atg | tcc | acc | ttc | gaa | ggg | ccg | atg | atg | gcc | atc | aag | acc | gtc | aac | 6076 |
| 2011 | Met | Ser | Thr | Phe | Glu | Gly | Pro | Met | Met | Ala | Ile | Lys | Thr | Val | Asn | 2025 |

| | | |
|------|---|------|
| 6077 | gcc ttg tcc cac tac acc gac tgg acc atc ggt cac gtc cat tcc | 6121 |
| 2026 | Ala Leu Ser His Tyr Thr Asp Trp Thr Ile Gly His Val His Ser | 2040 |
| | | |
| 6122 | ggc gcc ctg ggc tgg gtc ggc ttc gtc acc atg ggg gcg atg tac | 6166 |
| 2041 | Gly Ala Leu Gly Trp Val Gly Phe Val Thr Met Gly Ala Met Tyr | 2055 |
| | | |
| 6167 | tac ctg att ccg cgt ctg tgc ggc aag acc cag atg tgg agc aag | 6211 |
| 2056 | Tyr Leu Ile Pro Arg Leu Ser Gly Lys Thr Gln Met Trp Ser Lys | 2070 |
| | | |
| 6212 | gac ttg atc gag atc cac ttc tgg gtc gcc acc atc ggc atc gtg | 6256 |
| 2071 | Asp Leu Ile Glu Ile His Phe Trp Val Ala Thr Ile Gly Ile Val | 2085 |
| | | |
| 6257 | ctc tac atc gcc gcg atg tgg atc gcc ggt gta atg caa ggc ctg | 6301 |
| 2086 | Leu Tyr Ile Ala Ala Met Trp Ile Ala Gly Val Met Gln Gly Leu | 2100 |
| | | |
| 6302 | atg tgg cgc gcg gtc aat gcc gac ggc acc ctg acc tat acc ttc | 6346 |
| 2101 | Met Trp Arg Ala Val Asn Ala Asp Gly Thr Leu Thr Tyr Thr Phe | 2115 |
| | | |
| 6347 | gtg gaa agc gtc aag gca acc tat ccg tac tac gtg atc cgt ctg | 6391 |
| 2116 | Val Glu Ser Val Lys Ala Thr Tyr Pro Tyr Tyr Val Ile Arg Leu | 2130 |
| | | |
| 6392 | atg ggc ggc ctg atg tac ctg tcc ggc atg ctg gtg atg gcc tac | 6436 |
| 2131 | Met Gly Gly Leu Met Tyr Leu Ser Gly Met Leu Val Met Ala Tyr | 2145 |
| | | |
| 6437 | aac acc tgg cgc acc atg cgc ggc acc gag ctg gct gtc gca ccg | 6481 |
| 2146 | Asn Thr Trp Arg Thr Met Arg Gly Thr Glu Leu Ala Val Ala Pro | 2160 |
| | | |
| 6482 | att cct gcc gtg acc aac gct gcg cac tga tt tgg ggg aaa atg | 6527 |
| 2161 | Ile Pro Ala Val Thr Asn Ala Ala His End Met | 2174 |
| | | |
| | | |
| 6528 | aag ttt tgc cat gaa tgg atc gag aag aac ccc tgg ctg ctg atc | 6572 |
| 2175 | Lys Phe Ser His Glu Trp Ile Glu Lys Asn Pro Trp Leu Leu Ile | 2189 |
| | | |
| 6573 | ggc ctg gtg ctg ctg gtg gtg agc atc ggc ggc ctg gcc gag atc | 6617 |
| 2190 | Gly Leu Val Leu Leu Val Val Ser Ile Gly Gly Leu Ala Glu Ile | 2204 |
| | | |
| 6618 | gtg ccg ctg ttc ttc cag aaa tgc acg acc gaa ccc atc gct ggc | 6662 |
| 2205 | Val Pro Leu Phe Phe Gln Lys Ser Thr Thr Glu Pro Ile Ala Gly | 2219 |
| | | |
| 6663 | ctg aag ccc tat tgc gcc ctg cgc ctg gct ggc cgc gac gtg tac | 6707 |
| 2220 | Leu Lys Pro Tyr Ser Ala Leu Arg Leu Ala Gly Arg Asp Val Tyr | 2234 |
| | | |
| 6708 | atc cgc gag ggt tgc tac aac tgc cac tgc cag atg atc cgc ccg | 6752 |
| 2235 | Ile Arg Glu Gly Cys Tyr Asn Cys His Ser Gln Met Ile Arg Pro | 2249 |
| | | |
| 6753 | ctg cgc gcc gag acc gag cgc tat ggc cac tat tcc gtg gct ggc | 6797 |
| 2250 | Leu Arg Ala Glu Thr Glu Arg Tyr Gly His Tyr Ser Val Ala Gly | 2264 |
| | | |
| 6798 | gaa tgc gtc tat gac cac ccc ttc cag tgg ggt tcc aag cgc acc | 6842 |
| 2265 | Glu Ser Val Tyr Asp His Pro Phe Gln Trp Gly Ser Lys Arg Thr | 2279 |
| | | |
| 6843 | ggt cct gac ctg gcc cgc gtg ggt acc cgc tac agc gac gaa tgg | 6887 |
| 2280 | Gly Pro Asp Leu Ala Arg Val Gly Thr Arg Tyr Ser Asp Glu Trp | 2294 |
| | | |
| 6888 | cat cgc gcc cac ctg cac aat ccg cgc gac gtg gtg ccc gag tcc | 6932 |
| 2295 | His Arg Ala His Leu His Asn Pro Arg Asp Val Val Pro Glu Ser | 2309 |
| | | |
| 6933 | aac atg ccg gcc tac cca tgg ctg gag cag gcc aag ctg gac aac | 6977 |
| 2310 | Asn Met Pro Ala Tyr Pro Trp Leu Glu Gln Ala Lys Leu Asp Asn | 2324 |
| | | |
| 6978 | tac gac atc gtc tgc cgc atg aag gcc ttg aag cgc atc ggc gac | 7022 |

| | | |
|------|--|------|
| 2325 | Tyr Asp Ile Val Ser Arg Met Lys Ala Leu Lys Arg Ile Gly Asp | 2339 |
| 7023 | ccg tac acc gaa gaa gac atc aag aat gcc ccg gcc gag ctg gtc | 7067 |
| 2340 | Pro Tyr Thr Glu Glu Asp Ile Lys Asn Ala Pro Ala Glu Leu Val | 2354 |
| 7068 | ggc aag acc gag cag gat gcg ctg atc gcc tac ctg caa ggc ctg | 7112 |
| 2355 | Gly Lys Thr Glu Gln Asp Ala Leu Ile Ala Tyr Leu Gln Gly Leu | 2369 |
| 7113 | ggc gtg ctg atc aag <u>gcg gag aga</u> taa g acc atg att gaa atg | 7155 |
| 2370 | Gly Val Leu Ile Lys Ala Glu Arg End Met | 2385 |
| 7156 | att acc gat cca cgc agc ctg atc acg ctg atc agc ttc atc acc | 7200 |
| 2386 | Ile Thr Asp Pro Arg Ser Leu Ile Thr Leu Ile Ser Phe Ile Thr | 2400 |
| 7201 | ttt gcc ggc atc ctg tgg tgg acc tac atc gcc cac aag ccc gct | 7245 |
| 2401 | Phe Ala Gly Ile Leu Trp Trp Thr Tyr Ile Gly His Lys Pro Ala | 2415 |
| 7246 | gac ttc aag caa gcc gag atg att ccc ttc gcc gat ggc gac atc | 7290 |
| 2416 | Asp Phe Lys Gln Ala Glu Met Ile Pro Phe Ala Asp Gly Asp Ile | 2430 |
| 7291 | ggc cag ccg gcc cag gaa ggc gtt gac <u>aag gag gtg</u> cgc cat ggc | 7335 |
| 2431 | Gly Gln Pro Ala Gln Glu Gly Val Asp Lys Glu Val Arg His Gly | 2445 |
| 7336 | tga ttc acc aat ggt ttc tgg aat atc tgg atc atc gtg ctg acc | 7381 |
| 2446 | End Phe Thr Asn Gly Phe Trp Asn Ile Trp Ile Ile Val Leu Thr | 2460 |
| 7382 | gtg ctg ggt atc gtc gcc tgc acg ctg ctg ctg tgg cag caa tcg | 7426 |
| 2461 | Val Leu Gly Ile Val Gly Cys Thr Leu Leu Leu Trp Gln Gln Ser | 2475 |
| 7427 | aac tgg aag gtc aag aag ggc gag gaa aag gcg gcc ggt tcc acc | 7471 |
| 2476 | Asn Trp Lys Val Lys Lys Gly Glu Glu Lys Ala Ala Gly Ser Thr | 2490 |
| 7472 | act ggc cac gtc tgg gac gaa gac ctg acc gag ctg aac aac ccg | 7516 |
| 2491 | Thr Gly His Val Trp Asp Glu Asp Leu Thr Glu Leu Asn Asn Pro | 2505 |
| 7517 | ctg ccg cgc tgg tgg atg tgg ctg ttc tac ctg acc atc ttc ttc | 7561 |
| 2506 | Leu Pro Arg Trp Trp Met Trp Leu Phe Tyr Leu Thr Ile Phe Phe | 2520 |
| 7562 | tcc atc ggc tac ctg gtg gcc tat ccg ggc ctg ggc aac ctg ccc | 7606 |
| 2521 | Ser Ile Gly Tyr Leu Val Ala Tyr Pro Gly Leu Gly Asn Leu Pro | 2535 |
| 7607 | ggc acc ctg ggc tgg cgt tcc acc ggc gag cac gat gcc gat gtg | 7651 |
| 2536 | Gly Thr Leu Gly Trp Arg Ser Thr Gly Glu His Asp Ala Asp Val | 2550 |
| 7652 | aag gcc gct gaa gcc aag tat ggt ccg ctg ttc aac cac tat ctc | 7696 |
| 2551 | Lys Ala Ala Glu Ala Lys Tyr Gly Pro Leu Phe Asn His Tyr Leu | 2565 |
| 7697 | aag caa gac ctg aag gtg gtg gct ggc gac tcc cag gcc cac gcc | 7741 |
| 2566 | Lys Gln Asp Leu Lys Val Val Ala Gly Asp Ser Gln Ala His Ala | 2580 |
| 7742 | atc ggc gag cgc ctg ttc ctg acc tac tgc gcc cag tgc cat ggt | 7786 |
| 2581 | Ile Gly Glu Arg Leu Phe Leu Thr Tyr Cys Ala Gln Cys His Gly | 2595 |
| 7787 | tcg gat gcc cgc ggc aac aag ggc ttc cct aac ctg acg gac aac | 7831 |
| 2596 | Ser Asp Ala Arg Gly Asn Lys Gly Phe Pro Asn Leu Thr Asp Asn | 2610 |
| 7832 | gac tgg ctg cat ggc ggc acc cct gaa atc atc aag gaa acc atc | 7876 |
| 2611 | Asp Trp Leu His Gly Gly Thr Pro Glu Ile Ile Lys Glu Thr Ile | 2625 |
| 7877 | atg aag ggc cgc cac ggc gtc atg ccg ccg atg gcc gca gcc gtg | 7921 |
| 2626 | Met Lys Gly Arg His Gly Val Met Pro Pro Met Ala Ala Ala Val | 2640 |
| 7922 | ggt acc gac aag gat gtc gac aac gtc gcc aac tac gtg ctc agc | 7966 |

| | | |
|------|---|------|
| 2641 | Gly Thr Asp Lys Asp Val Asp Asn Val Ala Asn Tyr Val Leu Ser | 2655 |
| 7967 | ctg tcc gat tcg acc cat gat ccg atc aag gcc gag ctg ggc aag | 8011 |
| 2656 | Leu Ser Asp Ser Thr His Asp Pro Ile Lys Ala Glu Leu Gly Lys | 2670 |
| 8012 | tcc aag ttc gct gcc tgc atg gcc tgc cac ggt ccg ggc ggc aag | 8056 |
| 2671 | Ser Lys Phe Ala Ala Cys Met Ala Cys His Gly Pro Gly Gly Lys | 2685 |
| 8057 | ggg aac cag gcc ctg ggc gcc ccc aat ctg tcg gac aag atc tgg | 8101 |
| 2686 | Gly Asn Gln Ala Leu Gly Ala Pro Asn Leu Ser Asp Lys Ile Trp | 2700 |
| 8102 | ctg tat ggc ggc agt atc gac acc atc aag gaa acc atc aac aag | 8146 |
| 2701 | Leu Tyr Gly Gly Ser Ile Asp Thr Ile Lys Glu Thr Ile Asn Lys | 2715 |
| 8147 | ggg cgc gac aat acc atg ccg gcg ttt ggc gag ttc ctc ggt gaa | 8191 |
| 2716 | Gly Arg Asp Asn Thr Met Pro Ala Phe Gly Glu Phe Leu Gly Glu | 2730 |
| 8192 | ccc aag gtg cat gtg ctg gct gcc tat gtg tgg agt ctt tcc aac | 8236 |
| 2731 | Pro Lys Val His Val Leu Ala Ala Tyr Val Trp Ser Leu Ser Asn | 2745 |
| 8237 | aag ccc agc aac gcg gca gcc aag taa ga cct cat gcc gcg cat | 8280 |
| 2746 | Lys Pro Ser Asn Ala Ala Ala Lys End | |
| 8281 | gag gtg cta gtt ccg ccg ccg ggg ccg acg ctg aca cag atc ggc | 8325 |
| 8326 | gcg gtc ccg gcg cag tag gta acg cgc tcc ctc cac aag acc agg | 8370 |
| 8371 | agc gca cag aga <u>gtg aag gca</u> gat cag gca aga cc atg gaa aaa gcc8417 Met: Glu Lys Ala2804 | |
| 8418 | gtg atg aaa ccc agg cag gac aag gaa ggg gac ccg cca ccc gac | 8462 |
| 2805 | Val Met Lys Pro Arg Gln Asp Lys Glu Gly Asp Pro Pro Pro Asp | 2819 |
| 8463 | acc ggt gat agc aag aca ggc aag gat gaa tat gcc gtc atc cgc | 8507 |
| 2820 | Thr Gly Asp Ser Lys Thr Gly Lys Asp Glu Tyr Ala Val Ile Arg | 2834 |
| 8508 | atg tat gcg gcg cgc gag cag atc tac ccg cgc gaa atc cag gga | 8552 |
| 2838 | Met Tyr Ala Ala Arg Glu Gln Ile Tyr Pro Arg Glu Ile Gln Gly | 2849 |
| 8553 | cgt ttc gcc agc ctg cgc tag ctt tgt gtc ttt ctg acg cag ctg | 8597 |
| 2850 | Arg Phe Ala Ser Leu Arg Trp Leu Cys Val Phe Leu Thr Gln Leu | 2864 |
| 8598 | gtg ttc tac gga ctg ccg tgg atc aac tgg aat gag cgt cag gcc | 8642 |
| 2865 | Val Phe Tyr Gly Leu Pro Trp Ile Asn Trp Asn Glu Arg Gln Ala | 2879 |
| 8643 | gtg ctg ttc gac ctg gct tcc cgc aaa ttc tat ctg ttc ggc ctg | 8687 |
| 2880 | Val Leu Phe Asp Leu Ala Ser Arg Lys Phe Tyr Leu Phe Gly Leu | 2894 |
| 8688 | gtg atc tgg ccg cag gac ttc atc tgg ctg gcc gcg ctg ctg atc | 8732 |
| 2895 | Val Leu Trp Pro Gln Asp Phe Ile Trp Leu Ala Ala Leu Leu Ile | 2909 |
| 8733 | atc tgc gcc ttc agc ctg ttc ttg ttt aag gcc gtg gct gcc cgc | 8777 |
| 2910 | Ile Cys Ala Phe Ser Leu Phe Leu Phe Thr Ala Val Ala Gly Arg | 2924 |
| 8778 | gtc tgg tgc ggc tat tcc tgt ccg cag aag gtc tat acc gaa atc | 8822 |
| 2925 | Val Trp Cys Gly Tyr Ser Cys Pro Gln Thr Val Tyr Thr Glu Ile | 2939 |
| 8823 | ttc atg tgg ata gag cgc cgc atc gaa gcc aat cgc agc gcc cgc | 8867 |
| 2940 | Phe Met Trp Ile Glu Arg Arg Ile Glu Gly Asn Arg Ser Ala Arg | 2954 |
| 8868 | atg cgc ctg gac cgc caa ccc tgg tgg ttc gac aag ctg tgg cgc | 8912 |
| 2955 | Met Arg Leu Asp Arg Gln Pro Trp Ser Phe Asp Lys Leu Trp Arg | 2969 |

| | | |
|------|---|------|
| 8913 | aag tcy gcc aag cac ctg gcc tag ggc gcg gtg gcc ttg tgg acc | 8957 |
| 2970 | Lys Ser Ala Lys His Leu Ala Trp Gly Ala Val Ala Leu Trp Thr | 2984 |
| 8958 | ggc atc agc ttc gtc ggc tat ttc tcy ccc atc cgc gac ctg ctg | 9002 |
| 2985 | Gly Ile Ser Phe Val Gly Tyr Phe Ser Pro Ile Arg Asp Leu Leu | 2999 |
| 9001 | ccc gag atc agc ggc ttt gcg ctg ggt ccg tgg gag agc ttc tgg | 9047 |
| 3000 | Pro Glu Ile Ser Gly Phe Ala Leu Gly Pro Trp Glu Ser Phe Trp | 3014 |
| 9048 | atc ctg ttc tac ggc ttc gcc acc tat ggt aat gcg ggc tgg atg | 9092 |
| 3015 | Ile Leu Phe Tyr Gly Phe Ala Thr Tyr Gly Asn Ala Gly Trp Met | 3029 |
| 9093 | cgc gag cag gtc tgc aag cac atg tgt ccg tac gcc cgc ttc cag | 9137 |
| 3030 | Arg Glu Gln Val Cys Lys Tyr Met Cys Pro Tyr Ala Arg Phe Gln | 3044 |
| 9138 | agc gcc atg ttc gat cgt gat tcy ctg atc atc acc tat gac gtg | 9182 |
| 3045 | Ser Ala Met Phe Asp Arg Asp Ser Leu Ile Ile Thr Tyr Asp Val | 3059 |
| 9183 | gcc cgt gcc gaa cct cgt atg ccc gcc gcc aag gcg gcc aag ctg | 9227 |
| 3060 | Ala Arg Gly Glu Pro Arg Met Pro Ala Ala Lys Ala Ala Lys Leu | 3074 |
| 9228 | gac acc gcc gcc cgt gcc ggc gac tgc atc gac tgc acg atg tgc | 9272 |
| 3075 | Asp Thr Gly Ala Arg Ala Gly Asp Cys Ile Asp Cys Thr Met Cys | 3089 |
| 9273 | ctg cag gtc tgc ccc acc ggc atc gat atc cgc cag ggc ctg caa | 9317 |
| 3090 | Val Gln Val Cys Pro Thr Gly Ile Asp Ile Arg Gln Gly Leu Gln | 3104 |
| 9318 | tac atg tgc atc ggt tgc gcc gcc tgt gtg gat gcc tgc gac agc | 9362 |
| 3109 | Tyr Met Cys Ile Gly Cys Ala Ala Cys Val Asp Ala Cys Asp Ser | 3119 |
| 9363 | gtg atg gac aac atc aag cgg ccg cgt gaa ctg att cgc tac tcc | 9407 |
| 3120 | Val Met Asp Lys Ile Lys Arg Pro Arg Gly Leu Ile Arg Tyr Ser | 3134 |
| 9408 | acc gag aat gcg gtc gag cag ggc ttc tcc acg gcg gag atc cgg | 9452 |
| 3135 | Thr Glu Asn Ala Val Glu Gln Gly Phe Ser Thr Ala Glu Ile Arg | 3149 |
| 9453 | cgt cgt ctg gtg cgc ccg cgc atc ttg atc tac gcc gcc atc ctg | 9497 |
| 3150 | Arg Arg Leu Val Arg Pro Arg Ile Leu Ile Tyr Gly Ala Ile Leu | 3164 |
| 9498 | ggc gcc gtc att gcc ctg ttc gcc ggt tcy ctg tag gta cgc acc | 9542 |
| 3165 | Gly Ala Val Ile Ala Leu Phe Ala Gly Ser Leu Trp Val Arg Thr | 3179 |
| 9543 | ccg ctg aag ctg gac gtg atc cgc gac cgt ggt tcy atc ggg cgc | 9587 |
| 3180 | Pro Leu Lys Leu Asp Val Ile Arg Asp Arg Gly Ser Met Gly Arg | 3194 |
| 9588 | gaa gtc gaa gag ggc atc atc gaa gac gtc tat cgc ctg cag atc | 9632 |
| 3195 | Glu Val Glu Glu Gly Ile Ile Glu Asn Val Tyr Arg Leu Gln Ile | 3209 |
| 9633 | atc aac acc gat gaa cgg ggg cac cgt tat atc gta cgc gcc aag | 9677 |
| 3210 | Ile Asn Thr Asp Glu Arg Gly His Arg Tyr Leu Val Arg Ala Lys | 3224 |
| 9678 | ggc ctg gcc gga ctg agc gtg gat ccg gcg act ccc atc gag gtg | 9722 |
| 3225 | Gly Leu Ala Gly Leu Ser Val Asp Pro Ala Thr Pro Ile Glu Val | 3239 |
| 9723 | gcg gct acg cag acg gtc tcy gtg ccg gtg cgc gtg cgc gcg ccg | 9767 |
| 3240 | Ala Ala Thr Gln Thr Val Ser Val Pro Val Arg Val Arg Ala Pro | 3254 |
| 9765 | cac ggt gcg gcc gaa gtc ggc tcc aac aag atc cgc att gaa ctg | 9812 |
| 3255 | His Gly Ala Gly Glu Val Gly Ser Asn Lys Ile Arg Ile Glu Leu | 3269 |
| 9813 | gag gcc gaa gac cag cct gcg ctg aat gtc agc gaa aaa gcg gtg | 9857 |
| 3270 | Glu Ala Glu Asp Gln Pro Ala Leu Asn Val Ser Glu Lys Ala Val | 3284 |


```

9858 att ctg gty ccc cgg dgc tga tcg gac ccg atc tca cct aaa gag 9902
3245 Phe Leu Val Pro Arg Arg End

9903 gaa atc atg cag acc atg tct ccc aag ttc gcc agc act gct gcc 9947
Met Ser Pro Lys Phe Ala Ser Thr Ala Ala 3314

9948 gct gtt ccc tgg tat cgc cat cgc tgg ccc tgg ctc ctg atg gcc 9992
3315 Ala Val Pro Trp Tyr Arg His Arg Trp Pro Trp Leu Leu Met Ala 3329

9993 ggt ccg gcc gtg gtg gtg gtg gcg ggc gtc ttt acg gcc tgg ctg 10037
3330 Gly Pro Ala Val Val Val Val Ala Gly Val Phe Thr Ala Trp Leu 3344

10038 gcc atc tcg cgc gcc gat gcg ctg gtg gcg gac gac tac tac aag 10082
3345 Ala Ile Ser Arg Ala Asp Ala Leu Val Ala Asp Asp Tyr Tyr Lys 3359

10083 cag gcc aag gcc atc aac aag gac ttg cgc cgc gat cac gag gcc 10127
3360 Gln Gly Lys Ala Ile Asn Lys Asp Leu Arg Arg Asp His Glu Ala 3374

10128 cag cgc cta gcc gcg gcc atc gac atc gcc tat gac ccg gcg gcc 10172
3375 Gln Arg Leu Gly Ala Ala Ile Asp Ile Ala Tyr Asp Pro Ala Ala 3389

10173 ggc gta ctg cgt gcc cgc ctg caa ctg cgc cag ccg gcc ccg cag 10217
3390 Gly Val Leu Arg Gly Arg Leu Gln Leu Arg Gln Pro Ala Pro Gln 3404

10218 gcg cag tcc cag ctt gcc gat cag gca agc cgg aca ctg atc ctg 10262
3405 Ala Gln Ser Gln Leu Ala Asp Gln Ala Ser Arg Thr Leu Ile Leu 3419

10263 aac ctg gtg cac ccg acc cag ccg tcc aag gac cgt acg ctg cag 10307
3420 Asn Leu Val His Pro Thr Gln Pro Ser Lys Asp Arg Thr Leu Gln 3434

10308 gtg cag ccg gcc gcc gac gcc agt ttc tcg gtc ccg ctg gtg gaa 10352
3435 Val Gln Pro Ala Ala Asp Gly Ser Phe Ser Val Pro Leu Val Glu 3449

10353 atg gaa cag gcc cgc tgg cgt ctg gtg gcc gaa gac gcc agc cat 10397
3450 Met Glu Gln Ala Arg Trp Arg Leu Val Ala Glu Asp Gly Ser His 3464

10398 gtg tgg cgc ctg cac gcc acc tgg gcg tgg ccg caa cag cgc agc 10442
3465 Val Trp Arg Leu His Gly Thr Trp Ala Trp Pro Gln Gln Arg Ser 3479

10443 atg gaa atg aat gcc gag gcc tac gca ccg gcc gaa taa tgc gct 10487
3480 Met Glu Met Asn Ala Glu Ala Tyr Ala Pro Ala Glu End

10486 cta agc tga agc acc gag ccg acg caa aga tcg gca gtc tac gga 10530

10531 aga aat aac tga agg aga cag cca tgc gca aac agc aac gca aca 10575

10576 tcg cca tcg tga tct ggg tgg ctt tcc tga tgg cca cgg cgg ccg 10620

10621 agg gaa tat tct ttt cca tct tca acc ccg atg acc tgg cca tgg 10665

10666 cgg ccg gcc atg agt ggg agg cgc tgg gcg cct aca cgg tcg ggt 10710

10711 tct tct gtt tct ggg tga gtt ttg cga tct gtg gcc tgc tgg ccg 10755

10756 tga gcc tga cgc aga cct tgc cca ggg aac gtc tgc gtc acg tcg 10800

10801 ctt agc gcc cga tcg ctg ttg gcc gtt gac cgt tgg tcg ttg act 10845

10846 gcg gcc ccg ctg gcc gat gag gcc agc cct cag ctg gtg ctg cgg 10890
Val Leu Arg 3630

10891 gtc gcc gtc gcg gtg cag gcg tcc acg cct gcg gcc agg cgc ttc 10935

```

| | | |
|-------|---|-------|
| 3631 | Val Gly Val Ala Val Gln Ala Ser Thr Pro Ala Ala Arg Arg Phe | 3645 |
| 10936 | aat gcg gcc agg tgc act act tcc aca tca cgt tgc tcc acg gcc | 10980 |
| 3646 | Asn Ala Ala Arg Ser Thr Thr Ser Thr Ser Arg Cys Ser Thr Ala | 3660 |
| 10981 | agc aag ccc tgc ttg cgg aac ttg gag atc agc cgg ctg atg ctt | 11025 |
| 3661 | Ser Lys Pro Cys Leu Arg Asn Leu Glu Ile Ser Arg Leu Met Leu | 3675 |
| 11026 | tca atg gtc agg ccg agg taa ttg ccc acg tcc tgg cgt gtc atg | 11070 |
| 3676 | Ser Met Val Arg Pro Arg End Leu Pro Thr Ser Trp Arg Val Met | 3690 |
| 11071 | cgc agc tgg aaa cgc gtg gag gaa tag ccg cgt gcg gcg tag cgt | 11115 |
| 3691 | Arg Ser Trp Lys Arg Val Glu Glu End Pro Arg Ala Ala End Arg | 3705 |
| 11116 | gaa gag agg ttc acc agg aag gcg gca aaa cgc tgt tca gca cgc | 11160 |
| 3706 | Glu Glu Arg Phe Thr Arg Lys Ala Ala Lys Arg Cys Ser Ala Arg | 3720 |
| 11161 | atg ttc ccc agc agc atg atg acg ttc tgt tgc ctg gtg atc tgc | 11205 |
| 3721 | Met Phe Pro Ser Ser Met Met Thr Phe Cys Ser Leu Val Ile Ser | 3735 |
| 11206 | tgg ctc atg atg cga tgg aag tgg cgc agc aga tgg gga atc tgg | 11250 |
| 3736 | Trp Leu Met Met Arg Trp Lys Trp Arg Ser Arg Trp Gly Ile Trp | 3750 |
| 11251 | ccg aac aac tgt tcc agg cgg gcg aag ggg atc tgc cag act tgc | 11295 |
| 3751 | Pro Asn Asn Cys Ser Arg Arg Ala Lys Gly Ile Ser Gln Thr Ser | 3765 |
| 11296 | ctg tcc tgc agg gcc acg gcg tgc cac tga tgc tgt tgc gtg ctg | 11340 |
| 3766 | Leu Ser Cys Arg Ala Thr Ala Ser His End Cys Cys Ser Val Leu | 3780 |
| 11341 | atc gca tcc atg ccc agc agt tgc cca ggc atc tgg aag ccg gtg | 11385 |
| 3781 | Ile Ala Ser Met Pro Ser Ser Ser Pro Gly Ile Trp Lys Pro Val | 3795 |
| 11386 | atc tgg cgg tgc ccg tcc agg ttt tcc tga tag gtc ttg aaa tgg | 11430 |
| 3796 | Ile Trp Arg Ser Pro Ser Arg Phe Ser End End Val Leu Lys Trp | 3810 |
| 11431 | ccc acg cgc acc gca tag agc gcc gtg aag cgg tca ccg atg cga | 11475 |
| 3811 | Pro Thr Arg Thr Ala End Ser Ala Val Lys Arg Ser Pro Met Arg | 3825 |
| 11476 | tag agg aaa tgc tgc cgg gcg acc ttg cgg cga ccg atg atc ttg | 11520 |
| 3826 | End Arg Lys Ser Ser Arg Ala Thr Leu Arg Arg Pro Met Ile Leu | 3840 |
| 11521 | tcc agg cgc ttc atg tgc gac tca tcc aga ccc atg ggc agg cac | 11565 |
| 3841 | Ser Arg Arg Phe Met Ser Asp Ser Ser Arg Pro Met Gly Arg His | 3855 |
| 11566 | aac tgg tgc atg ccg cag gca gtg cag ctg cgc agc gcc gag ctg | 11610 |
| 3856 | Asn Trp Cys Met | |
| 11611 | gcg gca gcc tgc cgg gcc gcg gtc agc gac gag cag tgg gtc agg | 11655 |
| 11656 | ggg gca gcg gcg tta gcc ggc gtg gtg atc gga gtg gag tgg caa | 11700 |
| 11701 | gac tgg ttc ata cga tgt tcc cgg cgg aca tgt ggc aat cga atg | 11745 |
| 11746 | act cga ttt taa gtg cgc gga aca ata tag gaa gca cag tat tcc | 11790 |
| 11791 | cga atc aga cgg cat ttc ttg tcc cag atc aaa cat gtc gga act | 11835 |
| 11836 | aaa cgg ccg gca acg cta cgg aag gcc gga aga ccg cca gat tga | 11880 |
| 11881 | ggg ggc aaa tac cac aca aat tgc atg caa gcg gac aga agg aga | 11925 |
| 11926 | ggg gct gca atg agg gta ggg gag cgg cct tga tgg ggg aaa gtg | 11970 |


```

11971 ggc ggt gtc gcg ggg tta ccg ggg acc cgc gat tgc cct tcc atg 12015
12016 cga ccg tct tgt tat tga aac ttc cgt ggc ggt gaa act tcg gtg 12060
12061 ctg agg acg ggc gtc ctc agc ggg gtc atg ctt cgg cca 12099

```

APÊNDICE 1 - Em vermelho **proteína hipotética** conservada, em verde o gene *fixI*, em azul marinho *fixS*, em rosa *fixN*, em lilás *fixO*, em laranja *fixQ*, em bordô *fixP*, em salmão *fixG*, em azul claro *fixH* e a sequência de *fnr* está representada em verde limão. As regiões sublinhadas são os possíveis sítios de ligação ribossomal.

